B90

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 September 2001 (07.09.2001)

PCT

(10) International Publication Number
WO 01/64834 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US01/04926

(22) International Filing Date: 26 February 2001 (26.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

09/515,126	28 February 2000 (28.02.2000)	US
09/577,409	18 May 2000 (18.05.2000)	US
09/597,707	17 June 2000 (17.06.2000)	US
09/616,807	14 July 2000 (14.07.2000)	US
09/664,641	19 September 2000 (19.09.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	09/616,807 (CIP)
Filed on	14 July 2000 (14.07.2000)
US	09/664,641 (CIP)
Filed on	19 September 2000 (19.09.2000)
US	09/515,126 (CIP)
Filed on	28 February 2000 (28.02.2000)
US	09/577,409 (CIP)
Filed on	18 May 2000 (18.05.2000)
US	09/597,707 (CIP)
Filed on	17 June 2000 (17.06.2000)

(71) Applicant *(for all designated States except US)*: **HYSEQ, INC.** [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **TANG, Y, Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **ZHOU, Ping** [CN/US]; 1461 Japaur Lane, San Jose, CA 95132 (US). **ASUNDI, Vinod** [US/US]; 709 Foster City Boulevard, Foster City, CA 94404 (US). **ZHANG, Jie** [CN/US]; 20800 Homestead Road #38B, Cupertino, CA 95014 (US). **ZHAO, Qing, A.** [CN/US]; Apt. B-210, 1028 South de Anza Boulevard, San Jose, CA 95129 (US). **REN, Feiyan** [CN/US]; 20685 Garden Manor Court, Cupertino, CA 95014 (US). **XUE, Aidong, J.** [CN/US]; 1621 South Mary Avenue, Sunnyvale, CA 94087 (US). **YANG, Yonghong** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **WEHRMAN, Tom** [US/US]; 300 Pasteur Drive, Edwards R314, Stanford U. Medical Center, Stanford, CA 94305 (US). **WANG, Jian-Rui** [CN/US]; 744 Stendahl Lane, Cupertino, CA 95014 (US). **MA, Yunqing** [CN/US]; 280 W. California Avenue, #206, Sunnyvale, CA 94086 (US). **WANG, Dunrui** [CN/US]; 932 La Palma Place, Milpitas, CA 95035 (US). **CHEN, Rui-hong** [US/US]; 1031 Flying Fish Street, Foster City, CA 94404 (US). **XU, Chongjun** [CN/US]; 2918 Manitoba Drive, San Jose, CA 95130 (US). **DRMANAC, Radoje** [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris,, Glovsky and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*
— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

WO 01/64834 A2

NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

1. TECHNICAL FIELD

The present invention provides novel polynucleotides and proteins encoded by such polynucleotides, along with uses for these polynucleotides and proteins, for example in therapeutic, diagnostic and research methods.

2. BACKGROUND

Technology aimed at the discovery of protein factors (including e.g., cytokines, such as lymphokines, interferons, CSFs, chemokines, and interleukins) has matured rapidly over the past decade. The now routine hybridization cloning and expression cloning techniques clone novel polynucleotides "directly" in the sense that they rely on information directly related to the discovered protein (i.e., partial DNA/amino acid sequence of the protein in the case of hybridization cloning; activity of the protein in the case of expression cloning). More recent "indirect" cloning techniques such as signal sequence cloning, which isolates DNA sequences based on the presence of a now well-recognized secretory leader sequence motif, as well as various PCR-based or low stringency hybridization-based cloning techniques, have advanced the state of the art by making available large numbers of DNA/amino acid sequences for proteins that are known to have biological activity, for example, by virtue of their secreted nature in the case of leader sequence cloning, by virtue of their cell or tissue source in the case of PCR-based techniques, or by virtue of structural similarity to other genes of known biological activity.

Identified polynucleotide and polypeptide sequences have numerous applications in, for example, diagnostics, forensics, gene mapping; identification of mutations responsible for genetic disorders or other traits, to assess biodiversity, and to produce many other types of data and products dependent on DNA and amino acid sequences.

3. SUMMARY OF THE INVENTION

The compositions of the present invention include novel isolated polypeptides, novel isolated polynucleotides encoding such polypeptides, including recombinant DNA molecules, cloned genes or degenerate variants thereof, especially naturally occurring variants such as allelic variants, antisense polynucleotide molecules, and antibodies that specifically recognize one or more epitopes present on such polypeptides, as well as hybridomas producing such antibodies.

The compositions of the present invention additionally include vectors, including expression vectors, containing the polynucleotides of the invention, cells genetically engineered to contain such polynucleotides and cells genetically engineered to express such polynucleotides.

The present invention relates to a collection or library of at least one novel nucleic acid sequence assembled from expressed sequence tags (ESTs) isolated mainly by sequencing by hybridization (SBH), and in some cases, sequences obtained from one or more public databases. The invention relates also to the proteins encoded by such polynucleotides, along with therapeutic, diagnostic and research utilities for these polynucleotides and proteins. These nucleic acid sequences are designated as SEQ ID NO: 1-172, or 345-516. The polypeptides sequences are designated SEQ ID NO: 173-344, or 517-688. The nucleic acids and polypeptides are provided in the Sequence Listing. In the nucleic acids provided in the Sequence Listing, A is adenosine; C is cytosine; G is guanine; T is thymine; and N is any of the four bases. In the amino acids provided in the Sequence Listing, * corresponds to the stop codon.

The nucleic acid sequences of the present invention also include, nucleic acid sequences that hybridize to the complement of SEQ ID NO: 1-172, or 345-516 under stringent hybridization conditions; nucleic acid sequences which are allelic variants or species homologues of any of the nucleic acid sequences recited above, or nucleic acid sequences that encode a peptide comprising a specific domain or truncation of the peptides encoded by SEQ ID NO: 1-172, or 345-516. A polynucleotide comprising a nucleotide sequence having at least 90% identity to an identifying sequence of SEQ ID NO: 1-172, or 345-516 or a degenerate variant or fragment thereof. The identifying sequence can be 100 base pairs in length.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-172, or 345-516. The sequence information can be a segment of any one of SEQ ID NO: 1-172, or 345-516 that uniquely identifies or represents the sequence information of SEQ ID NO: 1-172, or 345-516.

A collection as used in this application can be a collection of only one polynucleotide. The collection of sequence information or identifying information of each sequence can be provided on a nucleic acid array. In one embodiment, segments of sequence information is provided on a nucleic acid array to detect the polynucleotide that contains the segment. The array can be designed to detect full-match or mismatch to the polynucleotide that contains the segment. The collection can also be provided in a computer-readable format.

This invention also includes the reverse or direct complement of any of the nucleic acid sequences recited above; cloning or expression vectors containing the nucleic acid sequences; and host cells or organisms transformed with these expression vectors. Nucleic acid sequences (or their reverse or direct complements) according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology, such as use as hybridization probes, use as primers for PCR, use in an array, use in computer-readable media, use in sequencing

full-length genes, use for chromosome and gene mapping, use in the recombinant production of protein, and use in the generation of anti-sense DNA or RNA, their chemical analogs and the like.

In a preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-172, or 345-516 or novel segments or parts of the nucleic acids of the invention are used as primers in expression assays that are well known in the art. In a particularly preferred embodiment, the nucleic acid sequences of SEQ ID NO: 1-172, or 345-516 or novel segments or parts of the nucleic acids provided herein are used in diagnostics for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The isolated polynucleotides of the invention include, but are not limited to, a polynucleotide comprising any one of the nucleotide sequences set forth in SEQ ID NO: 1-172, or 345-516 ; a polynucleotide comprising any of the full length protein coding sequences of SEQ ID NO: 1-172, or 345-516 ; and a polynucleotide comprising any of the nucleotide sequences of the mature protein coding sequences of SEQ ID NO: 1-172, or 345-516. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent hybridization conditions to (a) the complement of any one of the nucleotide sequences set forth in SEQ ID NO: 1-172, or 345-516; (b) a nucleotide sequence encoding any one of the amino acid sequences set forth in the Sequence Listing; (c) a polynucleotide which is an allelic variant of any polynucleotides recited above; (d) a polynucleotide which encodes a species homolog (e.g. orthologs) of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of any of the polypeptides comprising an amino acid sequence set forth in the Sequence Listing.

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising any of the amino acid sequences set forth in SEQ ID NO: 173-344, or 517-688; or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides with biological activity that are encoded by (a) any of the polynucleotides having a nucleotide sequence set forth in SEQ ID NO: 1-172, or 345-516; or (b) polynucleotides that hybridize to the complement of the polynucleotides of (a) under stringent hybridization conditions. Biologically or immunologically active variants of any of the polypeptide sequences in the Sequence Listing, and "substantial equivalents" thereof (e.g., with at least about 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% amino acid sequence identity) that preferably retain biological activity are also contemplated. The polypeptides of the invention may be wholly or partially chemically synthesized but are preferably produced by recombinant means using the genetically engineered cells (e.g. host cells) of the invention.

The invention also provides compositions comprising a polypeptide of the invention. Polypeptide compositions of the invention may further comprise an acceptable carrier, such as a hydrophilic, e.g., pharmaceutically acceptable, carrier.

The invention also provides host cells transformed or transfected with a polynucleotide of the invention.

The invention also relates to methods for producing a polypeptide of the invention comprising growing a culture of the host cells of the invention in a suitable culture medium under conditions permitting expression of the desired polypeptide, and purifying the polypeptide from the culture or from the host cells. Preferred embodiments include those in which the protein produced by such process is a mature form of the protein.

Polynucleotides according to the invention have numerous applications in a variety of techniques known to those skilled in the art of molecular biology. These techniques include use as hybridization probes, use as oligomers, or primers, for PCR, use for chromosome and gene mapping, use in the recombinant production of protein, and use in generation of anti-sense DNA or RNA, their chemical analogs and the like. For example, when the expression of an mRNA is largely restricted to a particular cell or tissue type, polynucleotides of the invention can be used as hybridization probes to detect the presence of the particular cell or tissue mRNA in a sample using, *e.g., in situ* hybridization.

In other exemplary embodiments, the polynucleotides are used in diagnostics as expressed sequence tags for identifying expressed genes or, as well known in the art and exemplified by Vollrath et al., Science 258:52-59 (1992), as expressed sequence tags for physical mapping of the human genome.

The polypeptides according to the invention can be used in a variety of conventional procedures and methods that are currently applied to other proteins. For example, a polypeptide of the invention can be used to generate an antibody that specifically binds the polypeptide. Such antibodies, particularly monoclonal antibodies, are useful for detecting or quantitating the polypeptide in tissue. The polypeptides of the invention can also be used as molecular weight markers, and as a food supplement.

Methods are also provided for preventing, treating, or ameliorating a medical condition which comprises the step of administering to a mammalian subject a therapeutically effective amount of a composition comprising a polypeptide of the present invention and a pharmaceutically acceptable carrier.

In particular, the polypeptides and polynucleotides of the invention can be utilized, for example, in methods for the prevention and/or treatment of disorders involving aberrant protein expression or biological activity.

The present invention further relates to methods for detecting the presence of the polynucleotides or polypeptides of the invention in a sample. Such methods can, for example, be utilized as part of prognostic and diagnostic evaluation of disorders as recited herein and for the identification of subjects exhibiting a predisposition to such conditions. The invention provides a method for detecting the polynucleotides of the invention in a sample, comprising contacting the sample with a compound that binds to and forms a complex with the polynucleotide of interest for a period sufficient to form the complex and under conditions sufficient to form a complex and detecting the complex such that if a complex is detected, the polynucleotide of interest is detected. The invention also provides a method for detecting the polypeptides of the invention in a sample comprising contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex and detecting the formation of the complex such that if a complex is formed, the polypeptide is detected.

The invention also provides kits comprising polynucleotide probes and/or monoclonal antibodies, and optionally quantitative standards, for carrying out methods of the invention. Furthermore, the invention provides methods for evaluating the efficacy of drugs, and monitoring the progress of patients, involved in clinical trials for the treatment of disorders as recited above.

The invention also provides methods for the identification of compounds that modulate (i.e., increase or decrease) the expression or activity of the polynucleotides and/or polypeptides of the invention. Such methods can be utilized, for example, for the identification of compounds that can ameliorate symptoms of disorders as recited herein. Such methods can include, but are not limited to, assays for identifying compounds and other substances that interact with (*e.g.*, bind to) the polypeptides of the invention. The invention provides a method for identifying a compound that binds to the polypeptides of the invention comprising contacting the compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and detecting the complex by detecting the reporter gene sequence expression such that if expression of the reporter gene is detected the compound the binds to a polypeptide of the invention is identified.

The methods of the invention also provide methods for treatment which involve the administration of the polynucleotides or polypeptides of the invention to individuals exhibiting symptoms or tendencies. In addition, the invention encompasses methods for treating diseases or disorders as recited herein comprising administering compounds and other substances that modulate the overall activity of thc target gene products. Compounds and other substances can

effect such modulation either on the level of target gene/protein expression or target protein activity.

The polypeptides of the present invention and the polynucleotides encoding them are also useful for the same functions known to one of skill in the art as the polypeptides and polynucleotides to which they have homology (set forth in Table 2); for which they have a signature region (as set forth in Table 3); or for which they have homology to a gene family (as set forth in Table 4). If no homology is set forth for a sequence, then the polypeptides and polynucleotides of the present invention are useful for a variety of applications, as described herein, including use in arrays for detection.

4. DETAILED DESCRIPTION OF THE INVENTION

4.1 DEFINITIONS

It must be noted that as used herein and in the appended claims, the singular forms "a", "an" and "the" include plural references unless the context clearly dictates otherwise.

The term "active" refers to those forms of the polypeptide which retain the biologic and/or immunologic activities of any naturally occurring polypeptide. According to the invention, the terms "biologically active" or "biological activity" refer to a protein or peptide having structural, regulatory or biochemical functions of a naturally occurring molecule. Likewise "immunologically active" or "immunological activity" refers to the capability of the natural, recombinant or synthetic polypeptide to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "activated cells" as used in this application are those cells which are engaged in extracellular or intracellular membrane trafficking, including the export of secretory or enzymatic molecules as part of a normal or disease process.

The terms "complementary" or "complementarity" refer to the natural binding of polynucleotides by base pairing. For example, the sequence 5'-AGT-3' binds to the complementary sequence 3'-TCA-5'. Complementarity between two single-stranded molecules may be "partial" such that only some of the nucleic acids bind or it may be "complete" such that total complementarity exists between the single stranded molecules. The degree of complementarity between the nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands.

The term "embryonic stem cells (ES)" refers to a cell that can give rise to many differentiated cell types in an embryo or an adult, including the germ cells. The term "germ line stem cells (GSCs)" refers to stem cells derived from primordial stem cells that provide a steady

and continuous source of germ cells for the production of gametes. The term "primordial germ cells (PGCs)" refers to a small population of cells set aside from other cell lineages particularly from the yolk sac, mesenteries, or gonadal ridges during embryogenesis that have the potential to differentiate into germ cells and other cells. PGCs are the source from which GSCs and ES cells are derived. The PGCs, the GSCs and the ES cells are capable of self-renewal. Thus these cells not only populate the germ line and give rise to a plurality of terminally differentiated cells that comprise the adult specialized organs, but are able to regenerate themselves.

The term "expression modulating fragment," EMF, means a series of nucleotides which modulates the expression of an operably linked ORF or another EMF.

As used herein, a sequence is said to "modulate the expression of an operably linked sequence" when the expression of the sequence is altered by the presence of the EMF. EMFs include, but are not limited to, promoters, and promoter modulating sequences (inducible elements). One class of EMFs are nucleic acid fragments which induce the expression of an operably linked ORF in response to a specific regulatory factor or physiological event.

The terms "nucleotide sequence" or "nucleic acid" or "polynucleotide" or "oligonculeotide" are used interchangeably and refer to a heteropolymer of nucleotides or the sequence of these nucleotides. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA) or to any DNA-like or RNA-like material. In the sequences herein A is adenine, C is cytosine, T is thymine, G is guanine and N is A, C, G or T (U). It is contemplated that where the polynucleotide is RNA, the T (thymine) in the sequences provided herein is substituted with U (uracil). Generally, nucleic acid segments provided by this invention may be assembled from fragments of the genome and short oligonucleotide linkers, or from a series of oligonucleotides, or from individual nucleotides, to provide a synthetic nucleic acid which is capable of being expressed in a recombinant transcriptional unit comprising regulatory elements derived from a microbial or viral operon, or a eukaryotic gene.

The terms "oligonucleotide fragment" or a "polynucleotide fragment", "portion," or "segment" or "probe" or "primer" are used interchangeably and refer to a sequence of nucleotide residues which are at least about 5 nucleotides, more preferably at least about 7 nucleotides, more preferably at least about 9 nucleotides, more preferably at least about 11 nucleotides and most preferably at least about 17 nucleotides. The fragment is preferably less than about 500 nucleotides, preferably less than about 200 nucleotides, more preferably less than about 100 nucleotides, more preferably less than about 50 nucleotides and most preferably less than 30 nucleotides. Preferably the probe is from about 6 nucleotides to about 200 nucleotides, preferably from about 15 to about 50 nucleotides, more preferably from about 17 to 30

nucleotides and most preferably from about 20 to 25 nucleotides. Preferably the fragments can be used in polymerase chain reaction (PCR), various hybridization procedures or microarray procedures to identify or amplify identical or related parts of mRNA or DNA molecules. A fragment or segment may uniquely identify each polynucleotide sequence of the present invention. Preferably the fragment comprises a sequence substantially similar to any one of SEQ ID NOs:1-20.

Probes may, for example, be used to determine whether specific mRNA molecules are present in a cell or tissue or to isolate similar nucleic acid sequences from chromosomal DNA as described by Walsh et al. (Walsh, P.S. et al., 1992, PCR Methods Appl 1:241-250). They may be labeled by nick translation, Klenow fill-in reaction, PCR, or other methods well known in the art. Probes of the present invention, their preparation and/or labeling are elaborated in Sambrook, J. et al., 1989, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY; or Ausubel, F.M. et al., 1989, Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, both of which are incorporated herein by reference in their cntirety.

The nucleic acid sequences of the present invention also include the sequence information from the nucleic acid sequences of SEQ ID NO: 1-172, or 345-516. The sequence information can be a segment of any one of SEQ ID NO: 1-172, or 345-516 that uniquely identifies or represents the sequence information of that sequence of SEQ ID NO: 1-172, or 345-516. One such segment can be a twenty-mer nucleic acid sequence because the probability that a twenty-mer is fully matched in the human genome is 1 in 300. In the human genome, there are three billion base pairs in one set of chromosomes. Because 4^{20} possible twenty-mers exist, there are 300 times more twenty-mers than there are base pairs in a set of human chromosomes. Using the same analysis, the probability for a seventeen-mer to be fully matched in the human genome is approximately 1 in 5. When these segments are used in arrays for expression studies, fifteen-mer segments can be used. The probability that the fifteen-mer is fully matched in the expressed sequences is also approximately one in five because expressed sequences comprise less than approximately 5% of the entire genome sequence.

Similarly, when using sequence information for detecting a single mismatch, a segment can be a twenty-five mer. The probability that the twenty-five mer would appear in a human genome with a single mismatch is calculated by multiplying the probability for a full match ($1 \div 4^{25}$) times the increased probability for mismatch at each nucleotide position (3 x 25). The probability that an eighteen mer with a single mismatch can be detected in an array for expression studies is approximately one in five. The probability that a twenty-mer with a single mismatch can be detected in a human genome is approximately one in five.

The term "open reading frame," ORF, means a series of nucleotide triplets coding for amino acids without any termination codons and is a sequence translatable into protein.

The terms “operably linked” or “operably associated” refer to functionally related nucleic acid sequences. For example, a promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the coding sequence. While operably linked nucleic acid sequences can be contiguous and in the same reading frame, certain genetic elements e.g. repressor genes are not contiguously linked to the coding sequence but still control transcription/translation of the coding sequence.

The term “pluripotent” refers to the capability of a cell to differentiate into a number of differentiated cell types that are present in an adult organism. A pluripotent cell is restricted in its differentiation capability in comparison to a totipotent cell.

The terms “polypeptide” or “peptide” or “amino acid sequence” refer to an oligopeptide, peptide, polypeptide or protein sequence or fragment thereof and to naturally occurring or synthetic molecules. A polypeptide "fragment," "portion," or "segment" is a stretch of amino acid residues of at least about 5 amino acids, preferably at least about 7 amino acids, more preferably at least about 9 amino acids and most preferably at least about 17 or more amino acids. The peptide preferably is not greater than about 500 amino acids, more preferably less than 200 amino acids more preferably less than 150 amino acids and most preferably less than 100 amino acids. Preferably the peptide is from about 5 to about 200 amino acids. To be active, any polypeptide must have sufficient length to display biological and/or immunological activity.

The term "naturally occurring polypeptide" refers to polypeptides produced by cells that have not been genetically engineered and specifically contemplates various polypeptides arising from post-translational modifications of the polypeptide including, but not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation.

The term “translated protein coding portion” means a sequence which encodes for the full length protein which may include any leader sequence or any processing sequence.

The term “mature protein coding sequence” means a sequence which encodes a peptide or protein without a signal or leader sequence. The “mature protein portion” means that portion of the protein which does not include a signal or leader sequence. The peptide may have been produced by processing in the cell which removes any leader/signal sequence. The mature protein portion may or may not include the initial methionine residue. The methionine residue may be removed from the protein during processing in the cell. The peptide may be produced synthetically or the protein may have been produced using a polynucleotide only encoding for the mature protein coding sequence.

The term "derivative" refers to polypeptides chemically modified by such techniques as ubiquitination, labeling (e.g., with radionuclides or various enzymes), covalent polymer attachment such as pegylation (derivatization with polyethylene glycol) and insertion or substitution by chemical synthesis of amino acids such as ornithine, which do not normally occur in human proteins.

The term "variant"(or "analog") refers to any polypeptide differing from naturally occurring polypeptides by amino acid insertions, deletions, and substitutions, created using, *e g.*, recombinant DNA techniques. Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequence.

Alternatively, recombinant variants encoding these same or similar polypeptides may be synthesized or selected by making use of the "redundancy" in the genetic code. Various codon substitutions, such as the silent changes which produce various restriction sites, may be introduced to optimize cloning into a plasmid or viral vector or expression in a particular prokaryotic or eukaryotic system. Mutations in the polynucleotide sequence may be reflected in the polypeptide or domains of other peptides added to the polypeptide to modify the properties of any part of the polypeptide, to change characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate.

Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, *i.e.*, conservative amino acid replacements. "Conservative" amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues involved. For example, nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan, and methionine; polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine; positively charged (basic) amino acids include arginine, lysine, and histidine; and negatively charged (acidic) amino acids include aspartic acid and glutamic acid. "Insertions" or "deletions" are preferably in the range of about 1 to 20 amino acids, more preferably 1 to 10 amino acids. The variation allowed may be experimentally determined by systematically making insertions, deletions, or substitutions of amino acids in a polypeptide molecule using recombinant DNA techniques and assaying the resulting recombinant variants for activity.

Alternatively, where alteration of function is desired, insertions, deletions or non-conservative alterations can be engineered to produce altered polypeptides. Such alterations

can, for example, alter one or more of the biological functions or biochemical characteristics of the polypeptides of the invention. For example, such alterations may change polypeptide characteristics such as ligand-binding affinities, interchain affinities, or degradation/turnover rate. Further, such alterations can be selected so as to generate polypeptides that are better suited for expression, scale up and the like in the host cells chosen for expression. For example, cysteine residues can be deleted or substituted with another amino acid residue in order to eliminate disulfide bridges.

The terms "purified" or "substantially purified" as used herein denotes that the indicated nucleic acid or polypeptide is present in the substantial absence of other biological macromolecules, *e.g.*, polynucleotides, proteins, and the like. In one embodiment, the polynucleotide or polypeptide is purified such that it constitutes at least 95% by weight, more preferably at least 99% by weight, of the indicated biological macromolecules present (but water, buffers, and other small molecules, especially molecules having a molecular weight of less than 1000 daltons, can be present).

The term "isolated" as used herein refers to a nucleic acid or polypeptide separated from at least one other component (e.g., nucleic acid or polypeptide) present with the nucleic acid or polypeptide in its natural source. In one embodiment, the nucleic acid or polypeptide is found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same. The terms "isolated" and "purified" do not encompass nucleic acids or polypeptides present in their natural source.

The term "recombinant," when used herein to refer to a polypeptide or protein, means that a polypeptide or protein is derived from recombinant (*e.g.*, microbial, insect, or mammalian) expression systems. "Microbial" refers to recombinant polypeptides or proteins made in bacterial or fungal (e.g., yeast) expression systems. As a product, "recombinant microbial" defines a polypeptide or protein essentially free of native endogenous substances and unaccompanied by associated native glycosylation. Polypeptides or proteins expressed in most bacterial cultures, *e.g.*, *E. coli*, will be free of glycosylation modifications; polypeptides or proteins expressed in yeast will have a glycosylation pattern in general different from those expressed in mammalian cells.

The term "recombinant expression vehicle or vector" refers to a plasmid or phage or virus or vector, for expressing a polypeptide from a DNA (RNA) sequence. An expression vehicle can comprise a transcriptional unit comprising an assembly of (1) a genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers, (2) a structural or coding sequence which is transcribed into mRNA and translated into protein, and (3) appropriate transcription initiation and termination sequences. Structural units intended for use

in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an amino terminal methionine residue. This residue may or may not be subsequently cleaved from the expressed recombinant protein to provide a final product.

The term "recombinant expression system" means host cells which have stably integrated a recombinant transcriptional unit into chromosomal DNA or carry the recombinant transcriptional unit extrachromosomally. Recombinant expression systems as defined herein will express heterologous polypeptides or proteins upon induction of the regulatory elements linked to the DNA segment or synthetic gene to be expressed. This term also means host cells which have stably integrated a recombinant genetic element or elements having a regulatory role in gene expression, for example, promoters or enhancers. Recombinant expression systems as defined herein will express polypeptides or proteins endogenous to the cell upon induction of the regulatory elements linked to the endogenous DNA segment or gene to be expressed. The cells can be prokaryotic or eukaryotic.

The term "secreted" includes a protein that is transported across or through a membrane, including transport as a result of signal sequences in its amino acid sequence when it is expressed in a suitable host cell. "Secreted" proteins include without limitation proteins secreted wholly (*e.g.*, soluble proteins) or partially (*e.g.*, receptors) from the cell in which they are expressed. "Secreted" proteins also include without limitation proteins that are transported across the membrane of the endoplasmic reticulum. "Secreted" proteins are also intended to include proteins containing non-typical signal sequences (e.g. Interleukin-1 Beta, see Krasney, P.A. and Young, P.R. (1992) Cytokine 4(2): 134 -143) and factors released from damaged cells (e.g. Interleukin-1 Receptor Antagonist, see Arend, W.P. et. al. (1998) Annu. Rev. Immunol. 16:27-55)

Where desired, an expression vector may be designed to contain a "signal or leader sequence" which will direct the polypeptide through the membrane of a cell. Such a sequence may be naturally present on the polypeptides of the present invention or provided from heterologous protein sources by recombinant DNA techniques.

The term "stringent" is used to refer to conditions that are commonly understood in the art as stringent. Stringent conditions can include highly stringent conditions (i.e., hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C), and moderately stringent conditions (i.e., washing in 0.2X SSC/0.1% SDS at 42°C). Other exemplary hybridization conditions are described herein in the examples.

In instances of hybridization of deoxyoligonucleotides, additional exemplary stringent hybridization conditions include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14-base oligonucleotides), 48°C (for 17-base oligos), 55°C (for 20-base oligonucleotides), and 60°C (for 23-base oligonucleotides).

As used herein, "substantially equivalent" can refer both to nucleotide and amino acid sequences, for example a mutant sequence, that varies from a reference sequence by one or more substitutions, deletions, or additions, the net effect of which does not result in an adverse functional dissimilarity between the reference and subject sequences. Typically, such a substantially equivalent sequence varies from one of those listed herein by no more than about 35% (*i.e.*, the number of individual residue substitutions, additions, and/or deletions in a substantially equivalent sequence, as compared to the corresponding reference sequence, divided by the total number of residues in the substantially equivalent sequence is about 0.35 or less). Such a sequence is said to have 65% sequence identity to the listed sequence. In one embodiment, a substantially equivalent, *e.g.*, mutant, sequence of the invention varies from a listed sequence by no more than 30% (70% sequence identity); in a variation of this embodiment, by no more than 25% (75% sequence identity); and in a further variation of this embodiment, by no more than 20% (80% sequence identity) and in a further variation of this embodiment, by no more than 10% (90% sequence identity) and in a further variation of this embodiment, by no more that 5% (95% sequence identity). Substantially equivalent, *e.g.*, mutant, amino acid sequences according to the invention preferably have at least 80% sequence identity with a listed amino acid sequence, more preferably at least 85% sequence identity, more preferably at least 90% sequence identity, more preferably at least 95% identity, more preferably at least 98% identity, and most preferably at least 99% identity. Substantially equivalent nucleotide sequences of the invention can have lower percent sequence identities, taking into account, for example, the redundancy or degeneracy of the genetic code. Preferably, nucleotide sequence has at least about 65% identity, more preferably at least about 75% identity, more preferably at least about 80% sequence identity, more preferably at least about 85% sequence identity, more preferably at least about 90% sequence identity, and most preferably at least about 95% identity, more preferably at least about 98% sequence identity, and most preferably at least about 99% sequence identity. For the purposes of the present invention, sequences having substantially equivalent biological activity and substantially equivalent expression characteristics are considered substantially equivalent. For the purposes of determining equivalence, truncation of the mature sequence (*e.g.*, via a mutation which creates a spurious stop codon) should be disregarded. Sequence identity may be determined, e.g., using the Jotun Hein method (Hein, J.

(1990) Methods Enzymol. 183:626-645). Identity between sequences can also be determined by other methods known in the art, e.g. by varying hybridization conditions.

The term "totipotent" refers to the capability of a cell to differentiate into all of the cell types of an adult organism.

The term "transformation" means introducing DNA into a suitable host cell so that the DNA is replicable, either as an extrachromosomal element, or by chromosomal integration. The term "transfection" refers to the taking up of an expression vector by a suitable host cell, whether or not any coding sequences are in fact expressed. The term "infection" refers to the introduction of nucleic acids into a suitable host cell by use of a virus or viral vector.

As used herein, an "uptake modulating fragment," UMF, means a series of nucleotides which mediate the uptake of a linked DNA fragment into a cell. UMFs can be readily identified using known UMFs as a target sequence or target motif with the computer-based systems described below. The presence and activity of a UMF can be confirmed by attaching the suspected UMF to a marker sequence. The resulting nucleic acid molecule is then incubated with an appropriate host under appropriate conditions and the uptake of the marker sequence is determined. As described above, a UMF will increase the frequency of uptake of a linked marker sequence.

Each of the above terms is meant to encompass all that is described for each, unless the context dictates otherwise.

4.2 NUCLEIC ACIDS OF THE INVENTION

Nucleotide sequences of the invention are set forth in the Sequence Listing.

The isolated polynucleotides of the invention include a polynucleotide comprising the nucleotide sequences of SEQ ID NO: 1-172, or 345-516; a polynucleotide encoding any one of the peptide sequences of SEQ ID NO: 173-344, or 517-688; and a polynucleotide comprising the nucleotide sequence encoding the mature protein coding sequence of the polypeptides of any one of SEQ ID NO: 173-344, or 517-688. The polynucleotides of the present invention also include, but are not limited to, a polynucleotide that hybridizes under stringent conditions to (a) the complement of any of the nucleotides sequences of SEQ ID NO: 1-172, or 345-516; (b) nucleotide sequences encoding any one of the amino acid sequences set forth in the Sequence Listing as SEQ ID NO: 173-344, or 517-688; (c) a polynucleotide which is an allelic variant of any polynucleotide recited above; (d) a polynucleotide which encodes a species homolog of any of the proteins recited above; or (e) a polynucleotide that encodes a polypeptide comprising a specific domain or truncation of the polypeptides of SEQ ID NO: 173-344, or 517-688. Domains of interest may depend on the nature of the encoded polypeptide; e.g., domains in

receptor-like polypeptides include ligand-binding, extracellular, transmembrane, or cytoplasmic domains, or combinations thereof; domains in immunoglobulin-like proteins include the variable immunoglobulin-like domains; domains in enzyme-like polypeptides include catalytic and substrate binding domains; and domains in ligand polypeptides include receptor-binding domains.

The polynucleotides of the invention include naturally occurring or wholly or partially synthetic DNA, e.g., cDNA and genomic DNA, and RNA, e.g., mRNA. The polynucleotides may include all of the coding region of the cDNA or may represent a portion of the coding region of the cDNA.

The present invention also provides genes corresponding to the cDNA sequences disclosed herein. The corresponding genes can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include the preparation of probes or primers from the disclosed sequence information for identification and/or amplification of genes in appropriate genomic libraries or other sources of genomic materials. Further 5' and 3' sequence can be obtained using methods known in the art. For example, full length cDNA or genomic DNA that corresponds to any of the polynucleotides of SEQ ID NO: 1-172, or 345-516 can be obtained by screening appropriate cDNA or genomic DNA libraries under suitable hybridization conditions using any of the polynucleotides of SEQ ID NO: 1-172, or 345-516 or a portion thereof as a probe. Alternatively, the polynucleotides of SEQ ID NO: 1-172, or 345-516 may be used as the basis for suitable primer(s) that allow identification and/or amplification of genes in appropriate genomic DNA or cDNA libraries.

The nucleic acid sequences of the invention can be assembled from ESTs and sequences (including cDNA and genomic sequences) obtained from one or more public databases, such as dbEST, gbpri, and UniGene. The EST sequences can provide identifying sequence information, representative fragment or segment information, or novel segment information for the full-length gene.

The polynucleotides of the invention also provide polynucleotides including nucleotide sequences that are substantially equivalent to the polynucleotides recited above. Polynucleotides according to the invention can have, *e.g.*, at least about 65%, at least about 70%, at least about 75%, at least about 80%, 81%, 82%, 83%, 84%, more typically at least about 85%, 86%, 87%, 88%, 89%, more typically at least about 90%, 91%, 92%, 93%, 94%, and even more typically at least about 95%, 96%, 97%, 98%, 99%, sequence identity to a polynucleotide recited above.

Included within the scope of the nucleic acid sequences of the invention are nucleic acid sequence fragments that hybridize under stringent conditions to any of the nucleotide sequences of SEQ ID NO: 1-172, or 345-516, or complements thereof, which fragment is greater than about

5 nucleotides, preferably 7 nucleotides, more preferably greater than 9 nucleotides and most preferably greater than 17 nucleotides. Fragments of, e.g. 15, 17, or 20 nucleotides or more that are selective for (i.e. specifically hybridize to any one of the polynucleotides of the invention) are contemplated. Probes capable of specifically hybridizing to a polynucleotide can differentiate polynucleotide sequences of the invention from other polynucleotide sequences in the same family of genes or can differentiate human genes from genes of other species, and are preferably based on unique nucleotide sequences.

The sequences falling within the scope of the present invention are not limited to these specific sequences, but also include allelic and species variations thereof. Allelic and species variations can be routinely determined by comparing the sequence provided SEQ ID NO: 1-172, or 345-516, a representative fragment thereof, or a nucleotide sequence at least 90% identical, preferably 95% identical, to SEQ ID NO: 1-172, or 345-516 with a sequence from another isolate of the same species. Furthermore, to accommodate codon variability, the invention includes nucleic acid molecules coding for the same amino acid sequences as do the specific ORFs disclosed herein. In other words, in the coding region of an ORF, substitution of one codon for another codon that encodes the same amino acid is expressly contemplated.

The nearest neighbor or homology result for the nucleic acids of the present invention, including SEQ ID NO: 1-172, or 345-516, can be obtained by searching a database using an algorithm or a program. Preferably, a BLAST which stands for Basic Local Alignment Search Tool is used to search for local sequence alignments (Altshul, S.F. J Mol. Evol. 36 290-300 (1993) and Altschul S.F. et al. J. Mol. Biol. 21:403-410 (1990)). Alternatively a FASTA version 3 search against Genpept, using Fastxy algorithm.

Species homologs (or orthologs) of the disclosed polynucleotides and proteins are also provided by the present invention. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source from the desired species.

The invention also encompasses allelic variants of the disclosed polynucleotides or proteins; that is, naturally-occurring alternative forms of the isolated polynucleotide which also encode proteins which are identical, homologous or related to that encoded by the polynucleotides.

The nucleic acid sequences of the invention are further directed to sequences which encode variants of the described nucleic acids. These amino acid sequence variants may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. Nucleic acids

encoding the amino acid sequence variants are preferably constructed by mutating the polynucleotide to encode an amino acid sequence that does not occur in nature. These nucleic acid alterations can be made at sites that differ in the nucleic acids from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, *e.g.*, by substituting first with conservative choices (*e.g.*, hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (*e.g.*, hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site. Amino acid sequence deletions generally range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells and sequences such as FLAG or poly-histidine sequences useful for purifying the expressed protein.

In a preferred method, polynucleotides encoding the novel amino acid sequences are changed via site-directed mutagenesis. This method uses oligonucleotide sequences to alter a polynucleotide to encode the desired amino acid variant, as well as sufficient adjacent nucleotides on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Edelman et al., *DNA* 2:183 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller and Smith, *Nucleic Acids Res.* 10:6487-6500 (1982). PCR may also be used to create amino acid sequence variants of the novel nucleic acids. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives a polynucleotide encoding the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., *Gene* 34:315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., supra, and *Current Protocols in Molecular Biology*, Ausubel et al. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent

amino acid sequence may be used in the practice of the invention for the cloning and expression of these novel nucleic acids. Such DNA sequences include those which are capable of hybridizing to the appropriate novel nucleic acid sequence under stringent conditions.

Polynucleotides encoding preferred polypeptide truncations of the invention can be used to generate polynucleotides encoding chimeric or fusion proteins comprising one or more domains of the invention and heterologous protein sequences.

The polynucleotides of the invention additionally include the complement of any of the polynucleotides recited above. The polynucleotide can be DNA (genomic, cDNA, amplified, or synthetic) or RNA. Methods and algorithms for obtaining such polynucleotides are well known to those of skill in the art and can include, for example, methods for determining hybridization conditions that can routinely isolate polynucleotides of the desired sequence identities.

In accordance with the invention, polynucleotide sequences comprising the mature protein coding sequences corresponding to any one of SEQ ID NO: 1-172, or 345-516, or functional equivalents thereof, may be used to generate recombinant DNA molecules that direct the expression of that nucleic acid, or a functional equivalent thereof, in appropriate host cells. Also included are the cDNA inserts of any of the clones identified herein.

A polynucleotide according to the invention can be joined to any of a variety of other nucleotide sequences by well-established recombinant DNA techniques (see Sambrook J et al. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, NY). Useful nucleotide sequences for joining to polynucleotides include an assortment of vectors, e.g., plasmids, cosmids, lambda phage derivatives, phagemids, and the like, that are well known in the art. Accordingly, the invention also provides a vector including a polynucleotide of the invention and a host cell containing the polynucleotide. In general, the vector contains an origin of replication functional in at least one organism, convenient restriction endonuclease sites, and a selectable marker for the host cell. Vectors according to the invention include expression vectors, replication vectors, probe generation vectors, and sequencing vectors. A host cell according to the invention can be a prokaryotic or eukaryotic cell and can be a unicellular organism or part of a multicellular organism.

The present invention further provides recombinant constructs comprising a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-172, or 345-516or a fragment thereof or any other polynucleotides of the invention. In one embodiment, the recombinant constructs of the present invention comprise a vector, such as a plasmid or viral vector, into which a nucleic acid having any of the nucleotide sequences of SEQ ID NO: 1-172, or 345-516 or a fragment thereof is inserted, in a forward or reverse orientation. In the case of a vector comprising one of the ORFs of the present invention, the vector may further comprise regulatory sequences,

including for example, a promoter, operably linked to the ORF. Large numbers of suitable vectors and promoters are known to those of skill in the art and are commercially available for generating the recombinant constructs of the present invention. The following vectors are provided by way of example. Bacterial: pBs, phagescript, PsiX174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLneo, pSV2cat, pOG44, PXTI, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia).

The isolated polynucleotide of the invention may be operably linked to an expression control sequence such as the pMT2 or pED expression vectors disclosed in Kaufman et al., *Nucleic Acids Res.* 19, 4485-4490 (1991), in order to produce the protein recombinantly. Many suitable expression control sequences are known in the art. General methods of expressing recombinant proteins are also known and are exemplified in R. Kaufman, *Methods in Enzymology* 185, 537-566 (1990). As defined herein "operably linked" means that the isolated polynucleotide of the invention and an expression control sequence are situated within a vector or cell in such a way that the protein is expressed by a host cell which has been transformed (transfected) with the ligated polynucleotide/expression control sequence.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, *e.g.*, the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), a-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an amino terminal identification peptide imparting desired characteristics, *e.g.*, stabilization or simplified purification of expressed recombinant product. Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination

signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*, although others may also be employed as a matter of choice.

As a representative but non-limiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM 1 (Promega Biotech, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced or derepressed by appropriate means (*e.g.*, temperature shift or chemical induction) and cells are cultured for an additional period. Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Polynucleotides of the invention can also be used to induce immune responses. For example, as described in Fan et al., *Nat. Biotech.* 17:870-872 (1999), incorporated herein by reference, nucleic acid sequences encoding a polypeptide may be used to generate antibodies against the encoded polypeptide following topical administration of naked plasmid DNA or following injection, and preferably intramuscular injection of the DNA. The nucleic acid sequences are preferably inserted in a recombinant expression vector and may be in the form of naked DNA.

4.3 ANTISENSE NUCLEIC ACIDS

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO: 1-172, or 345-516, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein, *e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a protein of any of

SEQ ID NO: 173-344, or 517-688 or antisense nucleic acids complementary to a nucleic acid sequence of SEQ ID NO: 1-172, or 345-516 are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence of the invention. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence of the invention. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding a nucleic acid disclosed herein (*e.g.*, SEQ ID NO: 1-172, or 345-516), antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of an mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of a mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of a mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used.

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the

antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a protein according to the invention to thereby inhibit expression of the protein, *e.g.*, by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, *e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier *et al.* (1987) *Nucleic Acids Res* 15: 6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue *et al.* (1987) *Nucleic Acids Res* 15: 6131-6148) or a chimeric RNA -DNA analogue (Inoue *et al.* (1987) *FEBS Lett* 215: 327-330).

4.4 RIBOZYMES AND PNA MOIETIES

In still another embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave a mRNA transcripts to thereby inhibit translation of a mRNA. A ribozyme having specificity for a nucleic acid of the invention can be

designed based upon the nucleotide sequence of a DNA disclosed herein (*i.e.*, SEQ ID NO: 1-172, or 345-516). For example, a derivative of a Tetrahymena L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a SECX-encoding mRNA. See, *e.g.*, Cech *et al.* U.S. Pat. No. 4,987,071; and Cech *et al.* U.S. Pat. No. 5,116,742. Alternatively, SECX mRNA can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. See, *e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region (*e.g.*, promoter and/or enhancers) to form triple helical structures that prevent transcription of the gene in target cells. See generally, Helene. (1991) *Anticancer Drug Des.* 6: 569-84; Helene. *et al.* (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14: 807-15.

In various embodiments, the nucleic acids of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup *et al.* (1996) *Bioorg Med Chem* 4: 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, *e.g.*, DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup *et al.* (1996) above; Perry-O'Keefe *et al.* (1996) *PNAS* 93: 14670-675.

PNAs of the invention can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of the invention can also be used, *e.g.*, in the analysis of single base pair mutations in a gene by, *e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S1 nucleases (Hyrup B. (1996) above); or as probes or primers for DNA sequence and hybridization (Hyrup *et al.* (1996), above; Perry-O'Keefe (1996), above).

In another embodiment, PNAs of the invention can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated that may

combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, *e.g.*, RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996) above). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996) above and Finn *et al.* (1996) *Nucl Acids Res* 24: 3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA (Mag *et al.* (1989) *Nucl Acid Res* 17: 5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn *et al.* (1996) above). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, Petersen *et al.* (1975) *Bioorg Med Chem Lett* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, *Proc. Natl. Acad. Sci. U.S.A.* 86:6553-6556; Lemaitre *et al.*, 1987, Proc. *Natl. Acad. Sci.* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, *e.g.*, PCT Publication No. W089/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (See, *e.g.*, Krol *et al.*, 1988, *BioTechniques* 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, *Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, etc.

4.5 HOSTS

The present invention further provides host cells genetically engineered to contain the polynucleotides of the invention. For example, such host cells may contain nucleic acids of the invention introduced into the host cell using known transformation, transfection or infection methods. The present invention still further provides host cells genetically engineered to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in the cell.

Knowledge of nucleic acid sequences allows for modification of cells to permit, or increase, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous

recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the polypeptide at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the encoding sequences. See, for example, PCT International Publication No. WO94/12650, PCT International Publication No. WO92/20808, and PCT International Publication No. WO91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

The host cell can be a higher eukaryotic host cell, such as a mammalian cell, a lower eukaryotic host cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the recombinant construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, or electroporation (Davis, L. et al., *Basic Methods in Molecular Biology* (1986)). The host cells containing one of the polynucleotides of the invention, can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF) or can be used to produce a heterologous protein under the control of the EMF.

Any host/vector system can be used to express one or more of the ORFs of the present invention. These include, but are not limited to, eukaryotic hosts such as HeLa cells, Cv-1 cell, COS cells, 293 cells, and Sf9 cells, as well as prokaryotic host such as *E. coli* and *B. subtilis*. The most preferred cells are those which do not normally express the particular polypeptide or protein or which expresses the polypeptide or protein at low natural level. Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., in Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, New York (1989), the disclosure of which is hereby incorporated by reference.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell 23:175 (1981). Other cell lines capable of expressing a compatible vector are, for example, the C127, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3

cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5′ flanking nontranscribed sequences. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early promoter, enhancer, splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements. Recombinant polypeptides and proteins produced in bacterial culture are usually isolated by initial extraction from cell pellets, followed by one or more salting-out, aqueous ion exchange or size exclusion chromatography steps. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps. Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents.

Alternatively, it may be possible to produce the protein in lower eukaryotes such as yeast or insects or in prokaryotes such as bacteria. Potentially suitable yeast strains include *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Kluyveromyces* strains, *Candida*, or any yeast strain capable of expressing heterologous proteins. Potentially suitable bacterial strains include *Escherichia coli*, *Bacillus subtilis*, *Salmonella typhimurium*, or any bacterial strain capable of expressing heterologous proteins. If the protein is made in yeast or bacteria, it may be necessary to modify the protein produced therein, for example by phosphorylation or glycosylation of the appropriate sites, in order to obtain the functional protein. Such covalent attachments may be accomplished using known chemical or enzymatic methods.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequence include polyadenylation signals, mRNA stability elements, splice

sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the host cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.6 POLYPEPTIDES OF THE INVENTION

The isolated polypeptides of the invention include, but are not limited to, a polypeptide comprising: the amino acid sequences set forth as any one of SEQ ID NO: 173-344, or 517-688 or an amino acid sequence encoded by any one of the nucleotide sequences SEQ ID NO: 1-172, or 345-516 or the corresponding full length or mature protein. Polypeptides of the invention also include polypeptides preferably with biological or immunological activity that are encoded by: (a) a polynucleotide having any one of thc nucleotide sequences set forth in SEQ ID NO: 1-172,

or 345-516 or (b) polynucleotides encoding any one of the amino acid sequences set forth as SEQ ID NO: 173-344, or 517-688 or (c) polynucleotides that hybridize to the complement of the polynucleotides of either (a) or (b) under stringent hybridization conditions. The invention also provides biologically active or immunologically active variants of any of the amino acid sequences set forth as SEQ ID NO: 173-344, or 517-688 or the corresponding full length or mature protein; and "substantial equivalents" thereof (*e.g.*, with at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, 86%, 87%, 88%, 89%, at least about 90%, 91%, 92%, 93%, 94%, typically at least about 95%, 96%, 97%, more typically at least about 98%, or most typically at least about 99% amino acid identity) that retain biological activity. Polypeptides encoded by allelic variants may have a similar, increased, or decreased activity compared to polypeptides comprising SEQ ID NO: 173-344, or 517-688.

Fragments of the proteins of the present invention which are capable of exhibiting biological activity are also encompassed by the present invention. Fragments of the protein may be in linear form or they may be cyclized using known methods, for example, as described in H. U. Saragovi, et al., Bio/Technology 10, 773-778 (1992) and in R. S. McDowell, et al., J. Amer. Chem. Soc. 114, 9245-9253 (1992), both of which are incorporated herein by reference. Such fragments may be fused to carrier molecules such as immunoglobulins for many purposes, including increasing the valency of protein binding sites.

The present invention also provides both full-length and mature forms (for example, without a signal sequence or precursor sequence) of the disclosed proteins. The protein coding sequence is identified in the sequence listing by translation of the disclosed nucleotide sequences. The mature form of such protein may be obtained by expression of a full-length polynucleotide in a suitable mammalian cell or other host cell. The sequence of the mature form of the protein is also determinable from the amino acid sequence of the full-length form. Where proteins of the present invention are membrane bound, soluble forms of the proteins are also provided. In such forms, part or all of the regions causing the proteins to be membrane bound are deleted so that the proteins are fully secreted from the cell in which they are expressed.

Protein compositions of the present invention may further comprise an acceptable carrier, such as a hydrophilic, *e.g.*, pharmaceutically acceptable, carrier.

The present invention further provides isolated polypeptides encoded by the nucleic acid fragments of the present invention or by degenerate variants of the nucleic acid fragments of the present invention. By "degenerate variant" is intended nucleotide fragments which differ from a nucleic acid fragment of the present invention (*e.g.*, an ORF) by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence. Preferred nucleic acid fragments of the present invention are the ORFs that encode proteins.

A variety of methodologies known in the art can be utilized to obtain any one of the isolated polypeptides or proteins of the present invention. At the simplest level, the amino acid sequence can be synthesized using commercially available peptide synthesizers. The synthetically-constructed protein sequences, by virtue of sharing primary, secondary or tertiary structural and/or conformational characteristics with proteins may possess biological properties in common therewith, including protein activity. This technique is particularly useful in producing small peptides and fragments of larger polypeptides. Fragments are useful, for example, in generating antibodies against the native polypeptide. Thus, they may be employed as biologically active or immunological substitutes for natural, purified proteins in screening of therapeutic compounds and in immunological processes for the development of antibodies.

The polypeptides and proteins of the present invention can alternatively be purified from cells which have been altered to express the desired polypeptide or protein. As used herein, a cell is said to be altered to express a desired polypeptide or protein when the cell, through genetic manipulation, is made to produce a polypeptide or protein which it normally does not produce or which the cell normally produces at a lower level. One skilled in the art can readily adapt procedures for introducing and expressing either recombinant or synthetic sequences into eukaryotic or prokaryotic cells in order to generate a cell which produces one of the polypeptides or proteins of the present invention.

The invention also relates to methods for producing a polypeptide comprising growing a culture of host cells of the invention in a suitable culture medium, and purifying the protein from the cells or the culture in which the cells are grown. For example, the methods of the invention include a process for producing a polypeptide in which a host cell containing a suitable expression vector that includes a polynucleotide of the invention is cultured under conditions that allow expression of the encoded polypeptide. The polypeptide can be recovered from the culture, conveniently from the culture medium, or from a lysate prepared from the host cells and further purified. Preferred embodiments include those in which the protein produced by such process is a full length or mature form of the protein.

In an alternative method, the polypeptide or protein is purified from bacterial cells which naturally produce the polypeptide or protein. One skilled in the art can readily follow known methods for isolating polypeptides and proteins in order to obtain one of the isolated polypeptides or proteins of the present invention. These include, but are not limited to, immunochromatography, HPLC, size-exclusion chromatography, ion-exchange chromatography, and immuno-affinity chromatography. See, *e.g.*, Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag (1994); Sambrook, et al., in Molecular Cloning: *A Laboratory Manual*; Ausubel et al., *Current Protocols in Molecular Biology*. Polypeptide fragments that

retain biological/immunological activity include fragments comprising greater than about 100 amino acids, or greater than about 200 amino acids, and fragments that encode specific protein domains.

The purified polypeptides can be used in *in vitro* binding assays which are well known in the art to identify molecules which bind to the polypeptides. These molecules include but are not limited to, for e.g., small molecules, molecules from combinatorial libraries, antibodies or other proteins. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

In addition, the peptides of the invention or molecules capable of binding to the peptides may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells. The toxin-binding molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for SEQ ID NO: 173-344, or 517-688.

The protein of the invention may also be expressed as a product of transgenic animals, *e.g.*, as a component of the milk of transgenic cows, goats, pigs, or sheep which are characterized by somatic or germ cells containing a nucleotide sequence encoding the protein.

The proteins provided herein also include proteins characterized by amino acid sequences similar to those of purified proteins but into which modification are naturally provided or deliberately engineered. For example, modifications, in the peptide or DNA sequence, can be made by those skilled in the art using known techniques. Modifications of interest in the protein sequences may include the alteration, substitution, replacement, insertion or deletion of a selected amino acid residue in the coding sequence. For example, one or more of the cysteine residues may be deleted or replaced with another amino acid to alter the conformation of the molecule. Techniques for such alteration, substitution, replacement, insertion or deletion are well known to those skilled in the art (see, *e.g.*, U.S. Pat. No. 4,518,584). Preferably, such alteration, substitution, replacement, insertion or deletion retains the desired activity of the protein. Regions of the protein that are important for the protein function can be determined by various methods known in the art including the alanine-scanning method which involved systematic substitution of single or strings of amino acids with alanine, followed by testing the resulting alanine-containing variant for biological activity. This type of analysis determines the importance of the substituted amino acid(s) in biological activity. Regions of the protein that are important for protein function may be determined by the eMATRIX program.

Other fragments and derivatives of the sequences of proteins which would be expected to retain protein activity in whole or in part and are useful for screening or other immunological

methodologies may also be easily made by those skilled in the art given the disclosures herein. Such modifications are encompassed by the present invention.

The protein may also be produced by operably linking the isolated polynucleotide of the invention to suitable control sequences in one or more insect expression vectors, and employing an insect expression system. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *e.g.*, Invitrogen, San Diego, Calif., U.S.A. (the MaxBat™ kit), and such methods are well known in the art, as described in Summers and Smith, Texas Agricultural Experiment Station Bulletin No. 1555 (1987), incorporated herein by reference. As used herein, an insect cell capable of expressing a polynucleotide of the present invention is "transformed."

The protein of the invention may be prepared by culturing transformed host cells under culture conditions suitable to express the recombinant protein. The resulting expressed protein may then be purified from such culture (*i.e.*, from culture medium or cell extracts) using known purification processes, such as gel filtration and ion exchange chromatography. The purification of the protein may also include an affinity column containing agents which will bind to the protein; one or more column steps over such affinity resins as concanavalin A-agarose, heparin-toyopearl™ or Cibacrom blue 3GA Sepharose™; one or more steps involving hydrophobic interaction chromatography using such resins as phenyl ether, butyl ether, or propyl ether; or immunoaffinity chromatography.

Alternatively, the protein of the invention may also be expressed in a form which will facilitate purification. For example, it may be expressed as a fusion protein, such as those of maltose binding protein (MBP), glutathione-S-transferase (GST) or thioredoxin (TRX), or as a His tag. Kits for expression and purification of such fusion proteins are commercially available from New England BioLab (Beverly, Mass.), Pharmacia (Piscataway, N.J.) and Invitrogen, respectively. The protein can also be tagged with an epitope and subsequently purified by using a specific antibody directed to such epitope. One such epitope ("FLAG®") is commercially available from Kodak (New Haven, Conn.).

Finally, one or more reverse-phase high performance liquid chromatography (RP- HPLC) steps employing hydrophobic RP-HPLC media, *e.g.*, silica gel having pendant methyl or other aliphatic groups, can be employed to further purify the protein. Some or all of the foregoing purification steps, in various combinations, can also be employed to provide a substantially homogeneous isolated recombinant protein. The protein thus purified is substantially free of other mammalian proteins and is defined in accordance with the present invention as an "isolated protein."

The polypeptides of the invention include analogs (variants). This embraces fragments, as well as peptides in which one or more amino acids has been deleted, inserted, or substituted. Also, analogs of the polypeptides of the invention embrace fusions of the polypeptides or modifications of the polypeptides of the invention, wherein the polypeptide or analog is fused to another moiety or moieties, e.g., targeting moiety or another therapeutic agent. Such analogs may exhibit improved properties such as activity and/or stability. Examples of moieties which may be fused to the polypeptide or an analog include, for example, targeting moieties which provide for the delivery of polypeptide to pancreatic cells, e.g., antibodies to pancreatic cells, antibodies to immune cells such as T-cells, monocytes, dendritic cells, granulocytes, etc., as well as receptor and ligands expressed on pancreatic or immune cells. Other moieties which may be fused to the polypeptide include therapeutic agents which are used for treatment, for example, immunosuppressive drugs such as cyclosporin, SK506, azathioprine, CD3 antibodies and steroids. Also, polypeptides may be fused to immune modulators, and other cytokines such as alpha or beta interferon.

4.6.1 DETERMINING POLYPEPTIDE AND POLYNUCLEOTIDE IDENTITY AND SIMILARITY

Preferred identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are codified in computer programs including, but are not limited to, the GCG program package, including GAP (Devereux, J., et al., Nucleic Acids Research 12(1):387 (1984); Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, BLASTX, FASTA (Altschul, S.F. et al., J. Molec. Biol. 215:403-410 (1990), PSI-BLAST (Altschul S.F. et al., Nucleic Acids Res. vol. 25, pp. 3389-3402, herein incorporated by reference), eMatrix software (Wu et al., J. Comp. Biol., Vol. 6, pp. 219-235 (1999), herein incorporated by reference), eMotif software (Nevill-Manning et al, ISMB-97, Vol. 4, pp. 202-209, herein incorporated by reference), pFam software (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1), pp. 320-322 (1998), herein incorporated by reference) and the Kyte-Doolittle hydrophobocity prediction algorithm (J. Mol Biol, 157, pp. 105-31 (1982), incorporated herein by reference). The BLAST programs are publicly available from the National Center for Biotechnology Information (NCBI) and other sources (BLAST Manual, Altschul, S., et al. NCB NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215:403-410 (1990).

4.7 CHIMERIC AND FUSION PROTEINS

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises a polypeptide of the invention operatively linked to

another polypeptide. Within a fusion protein the polypeptide according to the invention can correspond to all or a portion of a protein according to the invention. In one embodiment, a fusion protein comprises at least one biologically active portion of a protein according to the invention. In another embodiment, a fusion protein comprises at least two biologically active portions of a protein according to the invention. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide according to the invention and the other polypeptide are fused in-frame to each other. The polypeptide can be fused to the N-terminus or C-terminus.

For example, in one embodiment a fusion protein comprises a polypeptide according to the invention operably linked to the extracellular domain of a second protein. In another embodiment, the fusion protein is a GST-fusion protein in which the polypeptide sequences of the invention are fused to the C-terminus of the GST (i.e., glutathione S-transferase) sequences.

In another embodiment, the fusion protein is an immunoglobulin fusion protein in which the polypeptide sequences according to the invention comprise one or more domains fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand and a protein of the invention on the surface of a cell, to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion proteins can be used to affect the bioavailability of a cognate ligand. Inhibition of the ligand/protein interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, *e,g.,* cancer as well as modulating (*e.g.*, promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies in a subject, to purify ligands, and in screening assays to identify molecules that inhibit the interaction of a polypeptide of the invention with a ligand.

A chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently bc annealed and reamplificd to generate a chimeric genc sequence (scc, for

example, Ausubel et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the protein of the invention.

4.8 GENE THERAPY

Mutations in the polynucleotides of the invention gene may result in loss of normal function of the encoded protein. The invention thus provides gene therapy to restore normal activity of the polypeptides of the invention; or to treat disease states involving polypeptides of the invention. Delivery of a functional gene encoding polypeptides of the invention to appropriate cells is effected *ex vivo, in situ*, or *in vivo* by use of vectors, and more particularly viral vectors (e.g., adenovirus, adeno-associated virus, or a retrovirus), or *ex vivo* by use of physical DNA transfer methods (e.g., liposomes or chemical treatments). See, for example, Anderson, Nature, supplement to vol. 392, no. 6679, pp.25-20 (1998). For additional reviews of gene therapy technology see Friedmann, Science, 244: 1275-1281 (1989); Verma, Scientific American: 68-84 (1990); and Miller, Nature, 357: 455-460 (1992). Introduction of any one of the nucleotides of the present invention or a gene encoding the polypeptides of the present invention can also be accomplished with extrachromosomal substrates (transient expression) or artificial chromosomes (stable expression). Cells may also be cultured *ex vivo* in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced *in vivo* for therapeutic purposes. Alternatively, it is contemplated that in other human disease states, preventing the expression of or inhibiting the activity of polypeptides of the invention will be useful in treating the disease states. It is contemplated that antisense therapy or gene therapy could be applied to negatively regulate the expression of polypeptides of the invention.

Other methods inhibiting expression of a protein include the introduction of antisense molecules to the nucleic acids of the present invention, their complements, or their translated RNA sequences, by methods known in the art. Further, the polypeptides of the present invention can be inhibited by using targeted deletion methods, or the insertion of a negative regulatory element such as a silencer, which is tissue specific.

The present invention still further provides cells genetically engineered *in vivo* to express the polynucleotides of the invention, wherein such polynucleotides are in operative association with a regulatory sequence heterologous to the host cell which drives expression of the polynucleotides in

the cell. These methods can be used to increase or decrease the expression of the polynucleotides of the present invention.

Knowledge of DNA sequences provided by the invention allows for modification of cells to permit, increase, or decrease, expression of endogenous polypeptide. Cells can be modified (e.g., by homologous recombination) to provide increased polypeptide expression by replacing, in whole or in part, the naturally occurring promoter with all or part of a heterologous promoter so that the cells express the protein at higher levels. The heterologous promoter is inserted in such a manner that it is operatively linked to the desired protein encoding sequences. See, for example, PCT International Publication No. WO 94/12650, PCT International Publication No. WO 92/20808, and PCT International Publication No. WO 91/09955. It is also contemplated that, in addition to heterologous promoter DNA, amplifiable marker DNA (e.g., ada, dhfr, and the multifunctional CAD gene which encodes carbamyl phosphate synthase, aspartate transcarbamylase, and dihydroorotase) and/or intron DNA may be inserted along with the heterologous promoter DNA. If linked to the desired protein coding sequence, amplification of the marker DNA by standard selection methods results in co-amplification of the desired protein coding sequences in the cells.

In another embodiment of the present invention, cells and tissues may be engineered to express an endogenous gene comprising the polynucleotides of the invention under the control of inducible regulatory elements, in which case the regulatory sequences of the endogenous gene may be replaced by homologous recombination. As described herein, gene targeting can be used to replace a gene's existing regulatory region with a regulatory sequence isolated from a different gene or a novel regulatory sequence synthesized by genetic engineering methods. Such regulatory sequences may be comprised of promoters, enhancers, scaffold-attachment regions, negative regulatory elements, transcriptional initiation sites, regulatory protein binding sites or combinations of said sequences. Alternatively, sequences which affect the structure or stability of the RNA or protein produced may be replaced, removed, added, or otherwise modified by targeting. These sequences include polyadenylation signals, mRNA stability elements, splice sites, leader sequences for enhancing or modifying transport or secretion properties of the protein, or other sequences which alter or improve the function or stability of protein or RNA molecules.

The targeting event may be a simple insertion of the regulatory sequence, placing the gene under the control of the new regulatory sequence, *e.g.*, inserting a new promoter or enhancer or both upstream of a gene. Alternatively, the targeting event may be a simple deletion of a regulatory element, such as the deletion of a tissue-specific negative regulatory element. Alternatively, the targeting event may replace an existing element; for example, a tissue-specific enhancer can be replaced by an enhancer that has broader or different cell-type specificity than the naturally occurring elements. Here, the naturally occurring sequences are deleted and new sequences are

added. In all cases, the identification of the targeting event may be facilitated by the use of one or more selectable marker genes that are contiguous with the targeting DNA, allowing for the selection of cells in which the exogenous DNA has integrated into the cell genome. The identification of the targeting event may also be facilitated by the use of one or more marker genes exhibiting the property of negative selection, such that the negatively selectable marker is linked to the exogenous DNA, but configured such that the negatively selectable marker flanks the targeting sequence, and such that a correct homologous recombination event with sequences in the host cell genome does not result in the stable integration of the negatively selectable marker. Markers useful for this purpose include the Herpes Simplex Virus thymidine kinase (TK) gene or the bacterial xanthine-guanine phosphoribosyl-transferase (gpt) gene.

The gene targeting or gene activation techniques which can be used in accordance with this aspect of the invention are more particularly described in U.S. Patent No. 5,272,071 to Chappel; U.S. Patent No. 5,578,461 to Sherwin et al.; International Application No. PCT/US92/09627 (WO93/09222) by Selden et al.; and International Application No. PCT/US90/06436 (WO91/06667) by Skoultchi et al., each of which is incorporated by reference herein in its entirety.

4.9 TRANSGENIC ANIMALS

In preferred methods to determine biological functions of the polypeptides of the invention in vivo, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of a promoter of the polynucleotides of the invention is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous

promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

The polynucleotides of the present invention also make possible the development, through, e.g., homologous recombination or knock out strategies, of animals that fail to express polypeptides of the invention or that express a variant polypeptide. Such animals are useful as models for studying the *in vivo* activities of polypeptide as well as for studying modulators of the polypeptides of the invention.

In preferred methods to determine biological functions of the polypeptides of the invention *in vivo*, one or more genes provided by the invention are either over expressed or inactivated in the germ line of animals using homologous recombination [Capecchi, Science 244:1288-1292 (1989)]. Animals in which the gene is over expressed, under the regulatory control of exogenous or endogenous promoter elements, are known as transgenic animals. Animals in which an endogenous gene has been inactivated by homologous recombination are referred to as "knockout" animals. Knockout animals, preferably non-human mammals, can be prepared as described in U.S. Patent No. 5,557,032, incorporated herein by reference. Transgenic animals are useful to determine the roles polypeptides of the invention play in biological processes, and preferably in disease states. Transgenic animals are useful as model systems to identify compounds that modulate lipid metabolism. Transgenic animals, preferably non-human mammals, are produced using methods as described in U.S. Patent No 5,489,743 and PCT Publication No. WO94/28122, incorporated herein by reference.

Transgenic animals can be prepared wherein all or part of the polynucleotides of the invention promoter is either activated or inactivated to alter the level of expression of the polypeptides of the invention. Inactivation can be carried out using homologous recombination methods described above. Activation can be achieved by supplementing or even replacing the homologous promoter to provide for increased protein expression. The homologous promoter can be supplemented by insertion of one or more heterologous enhancer elements known to confer promoter activation in a particular tissue.

4.10 USES AND BIOLOGICAL ACTIVITY

The polynucleotides and proteins of the present invention are expected to exhibit one or more of the uses or biological activities (including those associated with assays cited herein) identified herein. Uses or activities described for proteins of the present invention may be provided by administration or use of such proteins or of polynucleotides encoding such proteins (such as, for example, in gene therapies or vectors suitable for introduction of DNA). The mechanism underlying the particular condition or pathology will dictate whether the

polypeptides of the invention, the polynucleotides of the invention or modulators (activators or inhibitors) thereof would be beneficial to the subject in need of treatment. Thus, "therapeutic compositions of the invention" include compositions comprising isolated polynucleotides (including recombinant DNA molecules, cloned genes and degenerate variants thereof) or polypeptides of the invention (including full length protein, mature protein and truncations or domains thereof), or compounds and other substances that modulate the overall activity of the target gene products, either at the level of target gene/protein expression or target protein activity. Such modulators include polypeptides, analogs, (variants), including fragments and fusion proteins, antibodies and other binding proteins; chemical compounds that directly or indirectly activate or inhibit the polypeptides of the invention (identified, e.g., via drug screening assays as described herein); antisense polynucleotides and polynucleotides suitable for triple helix formation; and in particular antibodies or other binding partners that specifically recognize one or more epitopes of the polypeptides of the invention.

The polypeptides of the present invention may likewise be involved in cellular activation or in one of the other physiological pathways described herein.

4.10.1 RESEARCH USES AND UTILITIES

The polynucleotides provided by the present invention can be used by the research community for various purposes. The polynucleotides can be used to express recombinant protein for analysis, characterization or therapeutic use; as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in disease states); as molecular weight markers on gels; as chromosome markers or tags (when labeled) to identify chromosomes or to map related gene positions; to compare with endogenous DNA sequences in patients to identify potential genetic disorders; as probes to hybridize and thus discover novel, related DNA sequences; as a source of information to derive PCR primers for genetic fingerprinting; as a probe to "subtract-out" known sequences in the process of discovering other novel polynucleotides; for selecting and making oligomers for attachment to a "gene chip" or other support, including for examination of expression patterns; to raise anti-protein antibodies using DNA immunization techniques; and as an antigen to raise anti-DNA antibodies or elicit another immune response. Where the polynucleotide encodes a protein which binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the polynucleotide can also be used in interaction trap assays (such as, for example, that described in Gyuris et al., Cell 75:791-803 (1993)) to identify polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction.

The polypeptides provided by the present invention can similarly be used in assays to determine biological activity, including in a panel of multiple proteins for high-throughput screening; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its receptor) in biological fluids; as markers for tissues in which the corresponding polypeptide is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state); and, of course, to isolate correlative receptors or ligands. Proteins involved in these binding interactions can also be used to screen for peptide or small molecule inhibitors or agonists of the binding interaction.

Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as research products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include without limitation "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

4.10.2 NUTRITIONAL USES

Polynucleotides and polypeptides of the present invention can also be used as nutritional sources or supplements. Such uses include without limitation use as a protein or amino acid supplement, use as a carbon source, use as a nitrogen source and use as a source of carbohydrate. In such cases the polypeptide or polynucleotide of the invention can be added to the feed of a particular organism or can be administered as a separate solid or liquid preparation, such as in the form of powder, pills, solutions, suspensions or capsules. In the case of microorganisms, the polypeptide or polynucleotide of the invention can be added to the medium in or on which the microorganism is cultured.

4.10.3 CYTOKINE AND CELL PROLIFERATION/DIFFERENTIATION ACTIVITY

A polypeptide of the present invention may exhibit activity relating to cytokine, cell proliferation (either inducing or inhibiting) or cell differentiation (either inducing or inhibiting) activity or may induce production of other cytokines in certain cell populations. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Many protein factors discovered to date, including all known cytokines, have exhibited activity in one or more factor-dependent cell proliferation assays, and hence the assays serve as a convenient

confirmation of cytokine activity. The activity of therapeutic compositions of the present invention is evidenced by any one of a number of routine factor dependent cell proliferation assays for cell lines including, without limitation, 32D, DA2, DA1G, T10, B9, B9/11, BaF3, MC9/G, M+(preB M+), 2E8, RB5, DA1, 123, T1165, HT2, CTLL2, TF-1, Mo7e, CMK, HUVEC, and Caco. Therapeutic compositions of the invention can be used in the following:

Assays for T-cell or thymocyte proliferation include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Bertagnolli et al., J. Immunol. 145:1706-1712, 1990; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Bertagnolli, et al., I. Immunol. 149:3778-3783, 1992; Bowman et al., I. Immunol. 152:1756-1761, 1994.

Assays for cytokine production and/or proliferation of spleen cells, lymph node cells or thymocytes include, without limitation, those described in: Polyclonal T cell stimulation, Kruisbeek, A. M. and Shevach, E. M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.12.1-3.12.14, John Wiley and Sons, Toronto. 1994; and Measurement of mouse and human interleukin-γ, Schreiber, R. D. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.8.1-6.8.8, John Wiley and Sons, Toronto. 1994.

Assays for proliferation and differentiation of hematopoietic and lymphopoietic cells include, without limitation, those described in: Measurement of Human and Murine Interleukin 2 and Interleukin 4, Bottomly, K., Davis, L. S. and Lipsky, P. E. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 6.3.1-6.3.12, John Wiley and Sons, Toronto. 1991; deVries et al., J. Exp. Med. 173:1205-1211, 1991; Moreau et al., Nature 336:690-692, 1988; Greenberger et al., Proc. Natl. Acad. Sci. U.S.A. 80:2931-2938, 1983; Measurement of mouse and human interleukin 6--Nordan, R. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.6.1-6.6.5, John Wiley and Sons, Toronto. 1991; Smith et al., Proc. Natl. Aced. Sci. U.S.A. 83:1857-1861, 1986; Measurement of human Interleukin 11--Bennett, F., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.15.1 John Wiley and Sons, Toronto. 1991; Measurement of mouse and human Interleukin 9--Ciarletta, A., Giannotti, J., Clark, S. C. and Turner, K. J. In Current Protocols in Immunology. J. E. Coligan eds. Vol 1 pp. 6.13.1, John Wiley and Sons, Toronto. 1991.

Assays for T-cell clone responses to antigens (which will identify, among others, proteins that affect APC-T cell interactions as well as direct T-cell effects by measuring proliferation and cytokine production) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W Strober,

Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, *In Vitro* assays for Mouse Lymphocyte Function; Chapter 6, Cytokines and their cellular receptors; Chapter 7, Immunologic studies in Humans); Weinberger et al., Proc. Natl. Acad. Sci. USA 77:6091-6095, 1980; Weinberger et al., Eur. J. Immun. 11:405-411, 1981; Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988.

4.10.4 STEM CELL GROWTH FACTOR ACTIVITY

A polypeptide of the present invention may exhibit stem cell growth factor activity and be involved in the proliferation, differentiation and survival of pluripotent and totipotent stem cells including primordial germ cells, embryonic stem cells, hematopoietic stem cells and/or germ line stem cells. Administration of the polypeptide of the invention to stem cells *in vivo* or *ex vivo* is expected to maintain and expand cell populations in a totipotential or pluripotential state which would be useful for re-engineering damaged or diseased tissues, transplantation, manufacture of bio-pharmaceuticals and the development of bio-sensors. The ability to produce large quantities of human cells has important working applications for the production of human proteins which currently must be obtained from non-human sources or donors, implantation of cells to treat diseases such as Parkinson's, Alzheimer's and other neurodegenerative diseases; tissues for grafting such as bone marrow, skin, cartilage, tendons, bone, muscle (including cardiac muscle), blood vessels, cornea, neural cells, gastrointestinal cells and others; and organs for transplantation such as kidney, liver, pancreas (including islet cells), heart and lung.

It is contemplated that multiple different exogenous growth factors and/or cytokines may be administered in combination with the polypeptide of the invention to achieve the desired effect, including any of the growth factors listed herein, other stem cell maintenance factors, and specifically including stem cell factor (SCF), leukemia inhibitory factor (LIF), Flt-3 ligand (Flt-3L), any of the interleukins, recombinant soluble IL-6 receptor fused to IL-6, macrophage inflammatory protein 1-alpha (MIP-1-alpha), G-CSF, GM-CSF, thrombopoietin (TPO), platelet factor 4 (PF-4), platelet-derived growth factor (PDGF), neural growth factors and basic fibroblast growth factor (bFGF).

Since totipotent stem cells can give rise to virtually any mature cell type, expansion of these cells in culture will facilitate the production of large quantities of mature cells. Techniques for culturing stem cells are known in the art and administration of polypeptides of the invention, optionally with other growth factors and/or cytokines, is expected to enhance the survival and proliferation of the stem cell populations. This can be accomplished by direct administration of the polypeptide of the invention to the culture medium. Alternatively, stroma cells transfected with a polynucleotide that encodes for the polypeptide of the invention can be used as a feeder

layer for the stem cell populations in culture or in vivo. Stromal support cells for feeder layers may include embryonic bone marrow fibroblasts, bone marrow stromal cells, fetal liver cells, or cultured embryonic fibroblasts (see U.S. Patent No. 5,690,926).

Stem cells themselves can be transfected with a polynucleotide of the invention to induce autocrine expression of the polypeptide of the invention. This will allow for generation of undifferentiated totipotential/pluripotential stem cell lines that are useful as is or that can then be differentiated into the desired mature cell types. These stable cell lines can also serve as a source of undifferentiated totipotential/pluripotential mRNA to create cDNA libraries and templates for polymerase chain reaction experiments. These studies would allow for the isolation and identification of differentially expressed genes in stem cell populations that regulate stem cell proliferation and/or maintenance.

Expansion and maintenance of totipotent stem cell populations will be useful in the treatment of many pathological conditions. For example, polypeptides of the present invention may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease, accidental damage or genetic disorders. The polypeptide of the invention may be useful for inducing the proliferation of neural cells and for the regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders which involve degeneration, death or trauma to neural cells or nerve tissue. In addition, the expanded stem cell populations can also be genetically altered for gene therapy purposes and to decrease host rejection of replacement tissues after grafting or implantation.

Expression of the polypeptide of the invention and its effect on stem cells can also be manipulated to achieve controlled differentiation of the stem cells into more differentiated cell types. A broadly applicable method of obtaining pure populations of a specific differentiated cell type from undifferentiated stem cell populations involves the use of a cell-type specific promoter driving a selectable marker. The selectable marker allows only cells of the desired type to survive. For example, stem cells can be induced to differentiate into cardiomyocytes (Wobus et al., Differentiation, 48: 173-182, (1991); Klug et al., J. Clin. Invest., 98(1): 216-224, (1998)) or skeletal muscle cells (Browder, L. W. In: *Principles of Tissue Engineering eds.* Lanza et al., Academic Press (1997)). Alternatively, directed differentiation of stem cells can be accomplished by culturing the stem cells in the presence of a differentiation factor such as retinoic acid and an antagonist of the polypeptide of the invention which would inhibit the effects of endogenous stem cell factor activity and allow differentiation to proceed.

In vitro cultures of stem cells can be used to determine if the polypeptide of the invention exhibits stem cell growth factor activity. Stem cells are isolated from any one of various cell

sources (including hematopoietic stem cells and embryonic stem cells) and cultured on a feeder layer, as described by Thompson et al. Proc. Natl. Acad. Sci, U.S.A., 92: 7844-7848 (1995), in the presence of the polypeptide of the invention alone or in combination with other growth factors or cytokines. The ability of the polypeptide of the invention to induce stem cells proliferation is determined by colony formation on semi-solid support e.g. as described by Bernstein et al., Blood, 77: 2316-2321 (1991).

4.10.5 HEMATOPOIESIS REGULATING ACTIVITY

A polypeptide of the present invention may be involved in regulation of hematopoiesis and, consequently, in the treatment of myeloid or lymphoid cell disorders. Even marginal biological activity in support of colony forming cells or of factor-dependent cell lines indicates involvement in regulating hematopoiesis, e.g. in supporting the growth and proliferation of erythroid progenitor cells alone or in combination with other cytokines, thereby indicating utility, for example, in treating various anemias or for use in conjunction with irradiation/chemotherapy to stimulate the production of erythroid precursors and/or erythroid cells; in supporting the growth and proliferation of myeloid cells such as granulocytes and monocytes/macrophages (i.e., traditional CSF activity) useful, for example, in conjunction with chemotherapy to prevent or treat consequent myelo-suppression; in supporting the growth and proliferation of megakaryocytes and consequently of platelets thereby allowing prevention or treatment of various platelet disorders such as thrombocytopenia, and generally for use in place of or complimentary to platelet transfusions; and/or in supporting the growth and proliferation of hematopoietic stem cells which are capable of maturing to any and all of the above-mentioned hematopoietic cells and therefore find therapeutic utility in various stem cell disorders (such as those usually treated with transplantation, including, without limitation, aplastic anemia and paroxysmal nocturnal hemoglobinuria), as well as in repopulating the stem cell compartment post irradiation/chemotherapy, either *in-vivo* or *ex-vivo* (i.e., in conjunction with bone marrow transplantation or with peripheral progenitor cell transplantation (homologous or heterologous)) as normal cells or genetically manipulated for gene therapy.

Therapeutic compositions of the invention can be used in the following:

Suitable assays for proliferation and differentiation of various hematopoietic lines are cited above.

Assays for embryonic stem cell differentiation (which will identify, among others, proteins that influence embryonic differentiation hematopoiesis) include, without limitation, those described in: Johansson et al. Cellular Biology 15:141-151, 1995; Keller et al., Molecular and Cellular Biology 13:473-486, 1993; McClanahan et al., Blood 81:2903-2915, 1993.

Assays for stem cell survival and differentiation (which will identify, among others, proteins that regulate lympho-hcmatopoiesis) include, without limitation, those described in: Methylcellulose colony forming assays, Freshney, M. G. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 265-268, Wiley-Liss, Inc., New York, N.Y. 1994; Hirayama et al., Proc. Natl. Acad. Sci. USA 89:5907-5911, 1992; Primitive hematopoietic colony forming cells with high proliferative potential, McNiece, I. K. and Briddell, R. A. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 23-39, Wiley-Liss, Inc., New York, N.Y. 1994; Neben et al., Experimental Hematology 22:353-359, 1994; Cobblestone area forming cell assay, Ploemacher, R. E. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 1-21, Wiley-Liss, Inc., New York, N.Y. 1994; Long term bone marrow cultures in the presence of stromal cells, Spooncer, E., Dexter, M. and Allen, T. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 163-179, Wiley-Liss, Inc., New York, N.Y. 1994; Long term culture initiating cell assay, Sutherland, H. J. In Culture of Hematopoietic Cells. R. I. Freshney, et al. eds. Vol pp. 139-162, Wiley-Liss, Inc., New York, N.Y. 1994.

4.10.6 TISSUE GROWTH ACTIVITY

A polypeptide of the present invention also may be involved in bone, cartilage, tendon, ligament and/or nerve tissue growth or regeneration, as well as in wound healing and tissue repair and replacement, and in healing of burns, incisions and ulcers.

A polypeptide of the present invention which induces cartilage and/or bone growth in circumstances where bone is not normally formed, has application in the healing of bone fractures and cartilage damage or defects in humans and other animals. Compositions of a polypeptide, antibody, binding partner, or other modulator of the invention may have prophylactic use in closed as well as open fracture reduction and also in the improved fixation of artificial joints. De novo bone formation induced by an osteogenic agent contributes to the repair of congenital, trauma induced, or oncologic resection induced craniofacial defects, and also is useful in cosmetic plastic surgery.

A polypeptide of this invention may also be involved in attracting bone-forming cells, stimulating growth of bone-forming cells, or inducing differentiation of progenitors of bone-forming cells. Treatment of osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease, such as through stimulation of bone and/or cartilage repair or by blocking inflammation or processes of tissue destruction (collagenase activity, osteoclast activity, etc.) mediated by inflammatory processes may also be possible using the composition of the invention.

Another category of tissue regeneration activity that may involve the polypeptide of the present invention is tendon/ligament formation. Induction of tendon/ligament-like tissue or other tissue formation in circumstances where such tissue is not normally formed, has application in the healing of tendon or ligament tears, deformities and other tendon or ligament defects in humans and other animals. Such a preparation employing a tendon/ligament-like tissue inducing protein may have prophylactic use in preventing damage to tendon or ligament tissue, as well as use in the improved fixation of tendon or ligament to bone or other tissues, and in repairing defects to tendon or ligament tissue. De novo tendon/ligament-like tissue formation induced by a composition of the present invention contributes to the repair of congenital, trauma induced, or other tendon or ligament defects of other origin, and is also useful in cosmetic plastic surgery for attachment or repair of tendons or ligaments. The compositions of the present invention may provide environment to attract tendon- or ligament-forming cells, stimulate growth of tendon- or ligament-forming cells, induce differentiation of progenitors of tendon- or ligament-forming cells, or induce growth of tendon/ligament cells or progenitors *ex vivo* for return *in vivo* to effect tissue repair. The compositions of the invention may also be useful in the treatment of tendinitis, carpal tunnel syndrome and other tendon or ligament defects. The compositions may also include an appropriate matrix and/or sequestering agent as a carrier as is well known in the art.

The compositions of the present invention may also be useful for proliferation of neural cells and for regeneration of nerve and brain tissue, i.e. for the treatment of central and peripheral nervous system diseases and neuropathies, as well as mechanical and traumatic disorders, which involve degeneration, death or trauma to neural cells or nerve tissue. More specifically, a composition may be used in the treatment of diseases of the peripheral nervous system, such as peripheral nerve injuries, peripheral neuropathy and localized neuropathies, and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome. Further conditions which may be treated in accordance with the present invention include mechanical and traumatic disorders, such as spinal cord disorders, head trauma and cerebrovascular diseases such as stroke. Peripheral neuropathies resulting from chemotherapy or other medical therapies may also be treatable using a composition of the invention.

Compositions of the invention may also be useful to promote better or faster closure of non-healing wounds, including without limitation pressure ulcers, ulcers associated with vascular insufficiency, surgical and traumatic wounds, and the like.

Compositions of the present invention may also be involved in the generation or regeneration of other tissues, such as organs (including, for example, pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac) and vascular (including vascular

endothelium) tissue, or for promoting the growth of cells comprising such tissues. Part of the desired effects may be by inhibition or modulation of fibrotic scarring may allow normal tissue to regenerate. A polypeptide of the present invention may also exhibit angiogenic activity.

A composition of the present invention may also be useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, and conditions resulting from systemic cytokine damage.

A composition of the present invention may also be useful for promoting or inhibiting differentiation of tissues described above from precursor tissues or cells; or for inhibiting the growth of tissues described above.

Therapeutic compositions of the invention can be used in the following:

Assays for tissue generation activity include, without limitation, those described in: International Patent Publication No. WO95/16035 (bone, cartilage, tendon); International Patent Publication No. WO95/05846 (nerve, neuronal); International Patent Publication No. WO91/07491 (skin, endothelium).

Assays for wound healing activity include, without limitation, those described in: Winter, Epidermal Wound Healing, pps. 71-112 (Maibach, H. I. and Rovee, D. T., eds.), Year Book Medical Publishers, Inc., Chicago, as modified by Eaglstein and Mertz, J. Invest. Dermatol 71:382-84 (1978).

4.10.7 IMMUNE STIMULATING OR SUPPRESSING ACTIVITY

A polypeptide of the present invention may also exhibit immune stimulating or immune suppressing activity, including without limitation the activities for which assays are described herein. A polynucleotide of the invention can encode a polypeptide exhibiting such activities. A protein may be useful in the treatment of various immune deficiencies and disorders (including severe combined immunodeficiency (SCID)), e.g., in regulating (up or down) growth and proliferation of T and/or B lymphocytes, as well as effecting the cytolytic activity of NK cells and other cell populations. These immune deficiencies may be genetic or be caused by viral (e.g., HIV) as well as bacterial or fungal infections, or may result from autoimmune disorders. More specifically, infectious diseases causes by viral, bacterial, fungal or other infection may be treatable using a protein of the present invention, including infections by HIV, hepatitis viruses, herpes viruses, mycobacteria, Leishmania spp., malaria spp. and various fungal infections such as candidiasis. Of course, in this regard, proteins of the present invention may also be useful where a boost to the immune system generally may be desirable, i.e., in the treatment of cancer.

Autoimmune disorders which may be treated using a protein of the present invention include, for example, connective tissue disease, multiple sclerosis, systemic lupus erythematosus,

rheumatoid arthritis, autoimmune pulmonary inflammation, Guillain-Barre syndrome, autoimmune thyroiditis, insulin dependent diabetes mellitis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. Such a protein (or antagonists thereof, including antibodies) of the present invention may also to be useful in the treatment of allergic reactions and conditions (*e.g.*, anaphylaxis, serum sickness, drug reactions, food allergies, insect venom allergies, mastocytosis, allergic rhinitis, hypersensitivity pneumonitis, urticaria, angioedema, eczema, atopic dermatitis, allergic contact dermatitis, erythema multiforme, Stevens-Johnson syndrome, allergic conjunctivitis, atopic keratoconjunctivitis, venereal keratoconjunctivitis, giant papillary conjunctivitis and contact allergies), such as asthma (particularly allergic asthma) or other respiratory problems. Other conditions, in which immune suppression is desired (including, for example, organ transplantation), may also be treatable using a protein (or antagonists thereof) of the present invention. The therapeutic effects of the polypeptides or antagonists thereof on allergic reactions can be evaluated by in vivo animals models such as the cumulative contact enhancement test (Lastbom et al., Toxicology 125: 59-66, 1998), skin prick test (Hoffmann et al., Allergy 54: 446-54, 1999), guinea pig skin sensitization test (Vohr et al., Arch. Toxocol. 73: 501-9), and murine local lymph node assay (Kimber et al., J. Toxicol. Environ. Health 53: 563-79).

Using the proteins of the invention it may also be possible to modulate immune responses, in a number of ways. Down regulation may be in the form of inhibiting or blocking an immune response already in progress or may involve preventing the induction of an immune response. The functions of activated T cells may be inhibited by suppressing T cell responses or by inducing specific tolerance in T cells, or both. Immunosuppression of T cell responses is generally an active, non-antigen-specific, process which requires continuous exposure of the T cells to the suppressive agent. Tolerance, which involves inducing non-responsiveness or anergy in T cells, is distinguishable from immunosuppression in that it is generally antigen-specific and persists after exposure to the tolerizing agent has ceased. Operationally, tolerance can be demonstrated by the lack of a T cell response upon reexposure to specific antigen in the absence of the tolerizing agent.

Down regulating or preventing one or more antigen functions (including without limitation B lymphocyte antigen functions (such as, for example, B7)), e.g., preventing high level lymphokine synthesis by activated T cells, will be useful in situations of tissue, skin and organ transplantation and in graft-versus-host disease (GVHD). For example, blockage of T cell function should result in reduced tissue destruction in tissue transplantation. Typically, in tissue transplants, rejection of the transplant is initiated through its recognition as foreign by T cells, followed by an immune reaction that destroys thc transplant. Thc administration of a therapeutic

Alternatively, anti-viral immune responses may be enhanced in an infected patient by removing T cells from the patient, costimulating the T cells in vitro with viral antigen-pulsed APCs either expressing a peptide of the present invention or together with a stimulatory form of a soluble peptide of the present invention and reintroducing the in vitro activated T cells into the patient. Another method of enhancing anti-viral immune responses would be to isolate infected cells from a patient, transfect them with a nucleic acid encoding a protein of the present invention as described herein such that the cells express all or a portion of the protein on their surface, and reintroduce the transfected cells into the patient. The infected cells would now be capable of delivering a costimulatory signal to, and thereby activate, T cells in vivo.

A polypeptide of the present invention may provide the necessary stimulation signal to T cells to induce a T cell mediated immune response against the transfected tumor cells. In addition, tumor cells which lack MHC class I or MHC class II molecules, or which fail to reexpress sufficient mounts of MHC class I or MHC class II molecules, can be transfected with nucleic acid encoding all or a portion of (e.g., a cytoplasmic-domain truncated portion) of an MHC class I alpha chain protein and β_2 microglobulin protein or an MHC class II alpha chain protein and an MHC class II beta chain protein to thereby express MHC class I or MHC class II proteins on the cell surface. Expression of the appropriate class I or class II MHC in conjunction with a peptide having the activity of a B lymphocyte antigen (e.g., B7-1, B7-2, B7-3) induces a T cell mediated immune response against the transfected tumor cell. Optionally, a gene encoding an antisense construct which blocks expression of an MHC class II associated protein, such as the invariant chain, can also be cotransfected with a DNA encoding a peptide having the activity of a B lymphocyte antigen to promote presentation of tumor associated antigens and induce tumor specific immunity. Thus, the induction of a T cell mediated immune response in a human subject may be sufficient to overcome tumor-specific tolerance in the subject.

The activity of a protein of the invention may, among other means, be measured by the following methods:

Suitable assays for thymocyte or splenocyte cytotoxicity include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Herrmann et al., Proc. Natl. Acad. Sci. USA 78:2488-2492, 1981; Herrmann et al., J. Immunol. 128:1968-1974, 1982; Handa et al., J. Immunol. 135:1564-1572, 1985; Takai et al., I. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bowman et al., J. Virology 61:1992-1998; Bertagnolli et al., Cellular Immunology 133:327-341, 1991; Brown et al., J. Immunol. 153:3079-3092, 1994.

Assays for T-cell-dependent immunoglobulin responses and isotype switching (which will identify, among others, proteins that modulate T-cell dependent antibody responses and that affect Th1/Th2 profiles) include, without limitation, those described in: Maliszewski, J. Immunol. 144:3028-3033, 1990; and Assays for B cell function: In vitro antibody production, Mond, J. J. and Brunswick, M. In Current Protocols in Immunology. J. E. e.a. Coligan eds. Vol 1 pp. 3.8.1-3.8.16, John Wiley and Sons, Toronto. 1994.

Mixed lymphocyte reaction (MLR) assays (which will identify, among others, proteins that generate predominantly Th1 and CTL responses) include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 3, In Vitro assays for Mouse Lymphocyte Function 3.1-3.19; Chapter 7, Immunologic studies in Humans); Takai et al., J. Immunol. 137:3494-3500, 1986; Takai et al., J. Immunol. 140:508-512, 1988; Bertagnolli et al., J. Immunol. 149:3778-3783, 1992.

Dendritic cell-dependent assays (which will identify, among others, proteins expressed by dendritic cells that activate naive T-cells) include, without limitation, those described in: Guery et al., J. Immunol. 134:536-544, 1995; Inaba et al., Journal of Experimental Medicine 173:549-559, 1991; Macatonia et al., Journal of Immunology 154:5071-5079, 1995; Porgador et al., Journal of Experimental Medicine 182:255-260, 1995; Nair et al., Journal of Virology 67:4062-4069, 1993; Huang et al., Science 264:961-965, 1994; Macatonia et al., Journal of Experimental Medicine 169:1255-1264, 1989; Bhardwaj et al., Journal of Clinical Investigation 94:797-807, 1994; and Inaba et al., Journal of Experimental Medicine 172:631-640, 1990.

Assays for lymphocyte survival/apoptosis (which will identify, among others, proteins that prevent apoptosis after superantigen induction and proteins that regulate lymphocyte homeostasis) include, without limitation, those described in: Darzynkiewicz et al., Cytometry 13:795-808, 1992; Gorczyca et al., Leukemia 7:659-670, 1993; Gorczyca et al., Cancer Research 53:1945-1951, 1993; Itoh et al., Cell 66:233-243, 1991; Zacharchuk, Journal of Immunology 145:4037-4045, 1990; Zamai et al., Cytometry 14:891-897, 1993; Gorczyca et al., International Journal of Oncology 1:639-648, 1992.

Assays for proteins that influence early steps of T-cell commitment and development include, without limitation, those described in: Antica et al., Blood 84:111-117, 1994; Fine et al., Cellular Immunology 155:111-122, 1994; Galy et al., Blood 85:2770-2778, 1995; Toki et al., Proc. Nat. Acad Sci. USA 88:7548-7551, 1991.

4.10.8 ACTIVIN/INHIBIN ACTIVITY

A polypeptide of the present invention may also exhibit activin- or inhibin-related activities. A polynucleotide of the invention may encode a polypeptide exhibiting such characteristics. Inhibins are characterized by their ability to inhibit the release of follicle stimulating hormone (FSH), while activins and are characterized by their ability to stimulate the release of follicle stimulating hormone (FSH). Thus, a polypeptide of the present invention, alone or in heterodimers with a member of the inhibin family, may be useful as a contraceptive based on the ability of inhibins to decrease fertility in female mammals and decrease spermatogenesis in male mammals. Administration of sufficient amounts of other inhibins can induce infertility in these mammals. Alternatively, the polypeptide of the invention, as a homodimer or as a heterodimer with other protein subunits of the inhibin group, may be useful as a fertility inducing therapeutic, based upon the ability of activin molecules in stimulating FSH release from cells of the anterior pituitary. See, for example, U.S. Pat. No. 4,798,885. A polypeptide of the invention may also be useful for advancement of the onset of fertility in sexually immature mammals, so as to increase the lifetime reproductive performance of domestic animals such as, but not limited to, cows, sheep and pigs.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods.

Assays for activin/inhibin activity include, without limitation, those described in: Vale et al., Endocrinology 91:562-572, 1972; Ling et al., Nature 321:779-782, 1986; Vale et al., Nature 321:776-779, 1986; Mason et al., Nature 318:659-663, 1985; Forage et al., Proc. Natl. Acad. Sci. USA 83:3091-3095, 1986.

4.10.9 CHEMOTACTIC/CHEMOKINETIC ACTIVITY

A polypeptide of the present invention may be involved in chemotactic or chemokinetic activity for mammalian cells, including, for example, monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Chemotactic and chemokinetic receptor activation can be used to mobilize or attract a desired cell population to a desired site of action. Chemotactic or chemokinetic compositions (e.g. proteins, antibodies, binding partners, or modulators of the invention) provide particular advantages in treatment of wounds and other trauma to tissues, as well as in treatment of localized infections. For example, attraction of lymphocytes, monocytes or neutrophils to tumors or sites of infection may result in improved immune responses against the tumor or infecting agent.

A protein or peptide has chemotactic activity for a particular cell population if it can stimulate, directly or indirectly, the directed orientation or movement of such cell population.

Preferably, the protein or peptide has the ability to directly stimulate directed movement of cells. Whether a particular protein has chemotactic activity for a population of cells can be readily determined by employing such protein or peptide in any known assay for cell chemotaxis.

Therapeutic compositions of the invention can be used in the following:

Assays for chemotactic activity (which will identify proteins that induce or prevent chemotaxis) consist of assays that measure the ability of a protein to induce the migration of cells across a membrane as well as the ability of a protein to induce the adhesion of one cell population to another cell population. Suitable assays for movement and adhesion include, without limitation, those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Marguiles, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley-Interscience (Chapter 6.12, Measurement of alpha and beta Chemokines 6.12.1-6.12.28; Taub et al. J. Clin. Invest. 95:1370-1376, 1995; Lind et al. APMIS 103:140-146, 1995; Muller et al Eur. J. Immunol. 25:1744-1748; Gruber et al. J. of Immunol. 152:5860-5867, 1994; Johnston et al. J. of Immunol. 153:1762-1768, 1994.

4.10.10 HEMOSTATIC AND THROMBOLYTIC ACTIVITY

A polypeptide of the invention may also be involved in hemostatis or thrombolysis or thrombosis. A polynucleotide of the invention can encode a polypeptide exhibiting such attributes. Compositions may be useful in treatment of various coagulation disorders (including hereditary disorders, such as hemophilias) or to enhance coagulation and other hemostatic events in treating wounds resulting from trauma, surgery or other causes. A composition of the invention may also be useful for dissolving or inhibiting formation of thromboses and for treatment and prevention of conditions resulting therefrom (such as, for example, infarction of cardiac and central nervous system vessels (e.g., stroke).

Therapeutic compositions of the invention can be used in the following:

Assay for hemostatic and thrombolytic activity include, without limitation, those described in: Linet et al., J. Clin. Pharmacol. 26:131-140, 1986; Burdick et al., Thrombosis Res. 45:413-419, 1987; Humphrey et al., Fibrinolysis 5:71-79 (1991); Schaub, Prostaglandins 35:467-474, 1988.

4.10.11 CANCER DIAGNOSIS AND THERAPY

Polypeptides of the invention may be involved in cancer cell generation, proliferation or metastasis. Detection of the presence or amount of polynucleotides or polypeptides of the invention may be useful for the diagnosis and/or prognosis of one or more types of cancer. For example, the presence or increased expression of a polynucleotide/polypeptide of the invention

may indicate a hereditary risk of cancer, a precancerous condition, or an ongoing malignancy. Conversely, a defect in the gene or absence of the polypeptide may be associated with a cancer condition. Identification of single nucleotide polymorphisms associated with cancer or a predisposition to cancer may also be useful for diagnosis or prognosis.

Cancer treatments promote tumor regression by inhibiting tumor cell proliferation, inhibiting angiogenesis (growth of new blood vessels that is necessary to support tumor growth) and/or prohibiting metastasis by reducing tumor cell motility or invasiveness. Therapeutic compositions of the invention may be effective in adult and pediatric oncology including in solid phase tumors/malignancies, locally advanced tumors, human soft tissue sarcomas, metastatic cancer, including lymphatic metastases, blood cell malignancies including multiple myeloma, acute and chronic leukemias, and lymphomas, head and neck cancers including mouth cancer, larynx cancer and thyroid cancer, lung cancers including small cell carcinoma and non-small cell cancers, breast cancers including small cell carcinoma and ductal carcinoma, gastrointestinal cancers including esophageal cancer, stomach cancer, colon cancer, colorectal cancer and polyps associated with colorectal neoplasia, pancreatic cancers, liver cancer, urologic cancers including bladder cancer and prostate cancer, malignancies of the female genital tract including ovarian carcinoma, uterine (including endometrial) cancers, and solid tumor in the ovarian follicle, kidney cancers including renal cell carcinoma, brain cancers including intrinsic brain tumors, neuroblastoma, astrocytic brain tumors, gliomas, metastatic tumor cell invasion in the central nervous system, bone cancers including osteomas, skin cancers including malignant melanoma, tumor progression of human skin keratinocytes, squamous cell carcinoma, basal cell carcinoma, hemangiopericytoma and Karposi's sarcoma.

Polypeptides, polynucleotides, or modulators of polypeptides of the invention (including inhibitors and stimulators of the biological activity of the polypeptide of the invention) may be administered to treat cancer. Therapeutic compositions can be administered in therapeutically effective dosages alone or in combination with adjuvant cancer therapy such as surgery, chemotherapy, radiotherapy, thermotherapy, and laser therapy, and may provide a beneficial effect, e.g. reducing tumor size, slowing rate of tumor growth, inhibiting metastasis, or otherwise improving overall clinical condition, without necessarily eradicating the cancer.

The composition can also be administered in therapeutically effective amounts as a portion of an anti-cancer cocktail. An anti-cancer cocktail is a mixture of the polypeptide or modulator of the invention with one or more anti-cancer drugs in addition to a pharmaceutically acceptable carrier for delivery. The use of anti-cancer cocktails as a cancer treatment is routine. Anti-cancer drugs that are well known in the art and can be used as a treatment in combination with the polypeptide or modulator of the invention include: Actinomycin D, Aminoglutethimide,

Asparaginase, Bleomycin, Busulfan, Carboplatin, Carmustine, Chlorambucil, Cisplatin (cis-DDP), Cyclophosphamide, Cytarabine HCl (Cytosine arabinoside), Dacarbazine, Dactinomycin, Daunorubicin HCl, Doxorubicin HCl, Estramustine phosphate sodium, Etoposide (V16-213), Floxuridine, 5-Fluorouracil (5-Fu), Flutamide, Hydroxyurea (hydroxycarbamide), Ifosfamide, Interferon Alpha-2a, Interferon Alpha-2b, Leuprolide acetate (LHRH-releasing factor analog), Lomustine, Mechlorethamine HCl (nitrogen mustard), Melphalan, Mercaptopurine, Mesna, Methotrexate (MTX), Mitomycin, Mitoxantrone HCl, Octreotide, Plicamycin, Procarbazine HCl, Streptozocin, Tamoxifen citrate, Thioguanine, Thiotepa, Vinblastine sulfate, Vincristine sulfate, Amsacrine, Azacitidine, Hexamethylmelamine, Interleukin-2, Mitoguazone, Pentostatin, Semustine, Teniposide, and Vindesine sulfate.

In addition, therapeutic compositions of the invention may be used for prophylactic treatment of cancer. There are hereditary conditions and/or environmental situations (e.g. exposure to carcinogens) known in the art that predispose an individual to developing cancers. Under these circumstances, it may be beneficial to treat these individuals with therapeutically effective doses of the polypeptide of the invention to reduce the risk of developing cancers.

In vitro models can be used to determine the effective doses of the polypeptide of the invention as a potential cancer treatment. These *in vitro* models include proliferation assays of cultured tumor cells, growth of cultured tumor cells in soft agar (see Freshney, (1987) Culture of Animal Cells: A Manual of Basic Technique, Wily-Liss, New York, NY Ch 18 and Ch 21), tumor systems in nude mice as described in Giovanella et al., J. Natl. Can. Inst., 52: 921-30 (1974), mobility and invasive potential of tumor cells in Boyden Chamber assays as described in Pilkington et al., Anticancer Res., 17: 4107-9 (1997), and angiogenesis assays such as induction of vascularization of the chick chorioallantoic membrane or induction of vascular endothelial cell migration as described in Ribatta et al., Intl. J. Dev. Biol., 40: 1189-97 (1999) and Li et al., Clin. Exp. Metastasis, 17:423-9 (1999), respectively. Suitable tumor cells lines are available, e.g. from American Type Tissue Culture Collection catalogs.

4.10.12 RECEPTOR/LIGAND ACTIVITY

A polypeptide of the present invention may also demonstrate activity as receptor, receptor ligand or inhibitor or agonist of receptor/ligand interactions. A polynucleotide of the invention can encode a polypeptide exhibiting such characteristics. Examples of such receptors and ligands include, without limitation, cytokine receptors and their ligands, receptor kinases and their ligands, receptor phosphatases and their ligands, receptors involved in cell-cell interactions and their ligands (including without limitation, cellular adhesion molecules (such as selectins, integrins and their ligands) and receptor/ligand pairs involved in antigen presentation, antigen

recognition and development of cellular and humoral immune responses. Receptors and ligands are also useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interaction. A protein of the present invention (including, without limitation, fragments of receptors and ligands) may themselves be useful as inhibitors of receptor/ligand interactions.

The activity of a polypeptide of the invention may, among other means, be measured by the following methods:

Suitable assays for receptor-ligand activity include without limitation those described in: Current Protocols in Immunology, Ed by J. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach, W. Strober, Pub. Greene Publishing Associates and Wiley- Interscience (Chapter 7.28, Measurement of Cellular Adhesion under static conditions 7.28.1- 7.28.22), Takai et al., Proc. Natl. Acad. Sci. USA 84:6864-6868, 1987; Bierer et al., J. Exp. Med. 168:1145-1156, 1988; Rosenstein et al., J. Exp. Med. 169:149-160 1989; Stoltenborg et al., J. Immunol. Methods 175:59-68, 1994; Stitt et al., Cell 80:661-670, 1995.

By way of example, the polypeptides of the invention may be used as a receptor for a ligand(s) thereby transmitting the biological activity of that ligand(s). Ligands may be identified through binding assays, affinity chromatography, dihybrid screening assays, BIAcore assays, gel overlay assays, or other methods known in the art.

Studies characterizing drugs or proteins as agonist or antagonist or partial agonists or a partial antagonist require the use of other proteins as competing ligands. The polypeptides of the present invention or ligand(s) thereof may be labeled by being coupled to radioisotopes, colorimetric molecules or a toxin molecules by conventional methods. ("Guide to Protein Purification" Murray P. Deutscher (ed) Methods in Enzymology Vol. 182 (1990) Academic Press, Inc. San Diego). Examples of radioisotopes include, but are not limited to, tritium and carbon-14 . Examples of colorimetric molecules include, but are not limited to, fluorescent molecules such as fluorescamine, or rhodamine or other colorimetric molecules. Examples of toxins include, but are not limited, to ricin.

4.10.13 DRUG SCREENING

This invention is particularly useful for screening chemical compounds by using the novel polypeptides or binding fragments thereof in any of a variety of drug screening techniques. The polypeptides or fragments employed in such a test may either be free in solution, affixed to a solid support, borne on a cell surface or located intracellularly. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant nucleic acids expressing the polypeptide or a fragment thereof. Drugs are screened against such

transformed cells in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may measure, for example, the formation of complexes between polypeptides of the invention or fragments and the agent being tested or examine the diminution in complex formation between the novel polypeptides and an appropriate cell line, which are well known in the art.

Sources for test compounds that may be screened for ability to bind to or modulate (i.e., increase or decrease) the activity of polypeptides of the invention include (1) inorganic and organic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of either random or mimetic peptides, oligonucleotides or organic molecules.

Chemical libraries may be readily synthesized or purchased from a number of commercial sources, and may include structural analogs of known compounds or compounds that are identified as "hits" or "leads" via natural product screening.

The sources of natural product libraries are microorganisms (including bacteria and fungi), animals, plants or other vegetation, or marine organisms, and libraries of mixtures for screening may be created by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of the organisms themselves. Natural product libraries include polyketides, non-ribosomal peptides, and (non-naturally occurring) variants thereof. For a review, see *Science 282*:63-68 (1998).

Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds and can be readily prepared by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, and polypeptide libraries. For a review of combinatorial chemistry and libraries created therefrom, see Myers, *Curr. Opin. Biotechnol.* 8:701-707 (1997). For reviews and examples of peptidomimetic libraries, see Al-Obeidi et al., *Mol. Biotechnol*, 9(3):205-23 (1998); Hruby et al., *Curr Opin Chem Biol*, 1(1):114-19 (1997); Dorner et al., *Bioorg Med Chem*, 4(5):709-15 (1996) (alkylated dipeptides).

Identification of modulators through use of the various libraries described herein permits modification of the candidate "hit" (or "lead") to optimize the capacity of the "hit" to bind a polypeptide of the invention. The molecules identified in the binding assay are then tested for antagonist or agonist activity in *in vivo* tissue culture or animal models that are well known in the art. In brief, the molecules are titrated into a plurality of cell cultures or animals and then tested for either cell/animal death or prolonged survival of the animal/cells.

The binding molecules thus identified may be complexed with toxins, e.g., ricin or cholera, or with other compounds that are toxic to cells such as radioisotopes. The toxin-binding

molecule complex is then targeted to a tumor or other cell by the specificity of the binding molecule for a polypeptide of the invention. Alternatively, the binding molecules may be complexed with imaging agents for targeting and imaging purposes.

4.10.14 ASSAY FOR RECEPTOR ACTIVITY

The invention also provides methods to detect specific binding of a polypeptide e.g. a ligand or a receptor. The art provides numerous assays particularly useful for identifying previously unknown binding partners for receptor polypeptides of the invention. For example, expression cloning using mammalian or bacterial cells, or dihybrid screening assays can be used to identify polynucleotides encoding binding partners. As another example, affinity chromatography with the appropriate immobilized polypeptide of the invention can be used to isolate polypeptides that recognize and bind polypeptides of the invention. There are a number of different libraries used for the identification of compounds, and in particular small molecules, that modulate (*i.e.,* increase or decrease) biological activity of a polypeptide of the invention. Ligands for receptor polypeptides of the invention can also be identified by adding exogenous ligands, or cocktails of ligands to two cells populations that are genetically identical except for the expression of the receptor of the invention: one cell population expresses the receptor of the invention whereas the other does not. The response of the two cell populations to the addition of ligands(s) are then compared. Alternatively, an expression library can be co-expressed with the polypeptide of the invention in cells and assayed for an autocrine response to identify potential ligand(s). As still another example, BIAcore assays, gel overlay assays, or other methods known in the art can be used to identify binding partner polypeptides, including, (1) organic and inorganic chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

The role of downstream intracellular signaling molecules in the signaling cascade of the polypeptide of the invention can be determined. For example, a chimeric protein in which the cytoplasmic domain of the polypeptide of the invention is fused to the extracellular portion of a protein, whose ligand has been identified, is produced in a host cell. The cell is then incubated with the ligand specific for the extracellular portion of the chimeric protein, thereby activating the chimeric receptor. Known downstream proteins involved in intracellular signaling can then be assayed for expected modifications i.e. phosphorylation. Other methods known to those in the art can also be used to identify signaling molecules involved in receptor activity.

4.10.15 ANTI-INFLAMMATORY ACTIVITY

Compositions of the present invention may also exhibit anti-inflammatory activity. The anti-inflammatory activity may be achieved by providing a stimulus to cells involved in the inflammatory response, by inhibiting or promoting cell-cell interactions (such as, for example, cell adhesion), by inhibiting or promoting chemotaxis of cells involved in the inflammatory process, inhibiting or promoting cell extravasation, or by stimulating or suppressing production of other factors which more directly inhibit or promote an inflammatory response. Compositions with such activities can be used to treat inflammatory conditions including chronic or acute conditions), including without limitation intimation associated with infection (such as septic shock, sepsis or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine-induced lung injury, inflammatory bowel disease, Crohn's disease or resulting from over production of cytokines such as TNF or IL-1. Compositions of the invention may also be useful to treat anaphylaxis and hypersensitivity to an antigenic substance or material. Compositions of this invention may be utilized to prevent or treat conditions such as, but not limited to, sepsis, acute pancreatitis, endotoxin shock, cytokine induced shock, rheumatoid arthritis, chronic inflammatory arthritis, pancreatic cell damage from diabetes mellitus type 1, graft versus host disease, inflammatory bowel disease, inflamation associated with pulmonary disease, other autoimmune disease or inflammatory disease, an antiproliferative agent such as for acute or chronic mylegenous leukemia or in the prevention of premature labor secondary to intrauterine infections.

4.10.16 LEUKEMIAS

Leukemias and related disorders may be treated or prevented by administration of a therapeutic that promotes or inhibits function of the polynucleotides and/or polypeptides of the invention. Such leukemias and related disorders include but are not limited to acute leukemia, acute lymphocytic leukemia, acute myelocytic leukemia, myeloblastic, promyelocytic, myelomonocytic, monocytic, erythroleukemia, chronic leukemia, chronic myelocytic (granulocytic) leukemia and chronic lymphocytic leukemia (for a review of such disorders, see Fishman et al., 1985, Medicine, 2d Ed., J.B. Lippincott Co., Philadelphia).

4.10.17 NERVOUS SYSTEM DISORDERS

Nervous system disorders, involving cell types which can be tested for efficacy of intervention with compounds that modulate the activity of the polynucleotides and/or polypeptides of the invention, and which can be treated upon thus observing an indication of therapeutic utility, include but are not limited to nervous system injuries, and diseases or

disorders which result in either a disconnection of axons, a diminution or degeneration of neurons, or demyelination. Nervous system lesions which may be treated in a patient (including human and non-human mammalian patients) according to the invention include but are not limited to the following lesions of either the central (including spinal cord, brain) or peripheral nervous systems:

(i) traumatic lesions, including lesions caused by physical injury or associated with surgery, for example, lesions which sever a portion of the nervous system, or compression injuries;

(ii) ischemic lesions, in which a lack of oxygen in a portion of the nervous system results in neuronal injury or death, including cerebral infarction or ischemia, or spinal cord infarction or ischemia;

(iii) infectious lesions, in which a portion of the nervous system is destroyed or injured as a result of infection, for example, by an abscess or associated with infection by human immunodeficiency virus, herpes zoster, or herpes simplex virus or with Lyme disease, tuberculosis, syphilis;

(iv) degenerative lesions, in which a portion of the nervous system is destroyed or injured as a result of a degenerative process including but not limited to degeneration associated with Parkinson's disease, Alzheimer's disease, Huntington's chorea, or amyotrophic lateral sclerosis;

(v) lesions associated with nutritional diseases or disorders, in which a portion of the nervous system is destroyed or injured by a nutritional disorder or disorder of metabolism including but not limited to, vitamin B12 deficiency, folic acid deficiency, Wernicke disease, tobacco-alcohol amblyopia, Marchiafava-Bignami disease (primary degeneration of the corpus callosum), and alcoholic cerebellar degeneration;

(vi) neurological lesions associated with systemic diseases including but not limited to diabetes (diabetic neuropathy, Bell's palsy), systemic lupus erythematosus, carcinoma, or sarcoidosis;

(vii) lesions caused by toxic substances including alcohol, lead, or particular neurotoxins; and

(viii) demyelinated lesions in which a portion of the nervous system is destroyed or injured by a demyelinating disease including but not limited to multiple sclerosis, human immunodeficiency virus-associated myelopathy, transverse myelopathy or various etiologies, progressive multifocal leukoencephalopathy, and central pontine myelinolysis.

Therapeutics which are useful according to the invention for treatment of a nervous system disorder may be selected by testing for biological activity in promoting the survival or

differentiation of neurons. For example, and not by way of limitation, therapeutics which elicit any of the following effects may be useful according to the invention:

(i) increased survival time of neurons in culture;

(ii) increased sprouting of neurons in culture or in vivo;

(iii) increased production of a neuron-associated molecule in culture or *in vivo*, *e.g.*, choline acetyltransferase or acetylcholinesterase with respect to motor neurons; or

(iv) decreased symptoms of neuron dysfunction in vivo.

Such effects may be measured by any method known in the art. In preferred, non-limiting embodiments, increased survival of neurons may be measured by the method set forth in Arakawa et al. (1990, J. Neurosci. 10:3507-3515); increased sprouting of neurons may be detected by methods set forth in Pestronk et al. (1980, Exp. Neurol. 70:65-82) or Brown et al. (1981, Ann. Rev. Neurosci. 4:17-42); increased production of neuron-associated molecules may be measured by bioassay, enzymatic assay, antibody binding, Northern blot assay, *etc.*, depending on the molecule to be measured; and motor neuron dysfunction may be measured by assessing the physical manifestation of motor neuron disorder, *e.g.*, weakness, motor neuron conduction velocity, or functional disability.

In specific embodiments, motor neuron disorders that may be treated according to the invention include but are not limited to disorders such as infarction, infection, exposure to toxin, trauma, surgical damage, degenerative disease or malignancy that may affect motor neurons as well as other components of the nervous system, as well as disorders that selectively affect neurons such as amyotrophic lateral sclerosis, and including but not limited to progressive spinal muscular atrophy, progressive bulbar palsy, primary lateral sclerosis, infantile and juvenile muscular atrophy, progressive bulbar paralysis of childhood (Fazio-Londe syndrome), poliomyelitis and the post polio syndrome, and Hereditary Motorsensory Neuropathy (Charcot-Marie-Tooth Disease).

4.10.18 OTHER ACTIVITIES

A polypeptide of the invention may also exhibit one or more of the following additional activities or effects: inhibiting the growth, infection or function of, or killing, infectious agents, including, without limitation, bacteria, viruses, fungi and other parasites; effecting (suppressing or enhancing) bodily characteristics, including, without limitation, height, weight, hair color, eye color, skin, fat to lean ratio or other tissue pigmentation, or organ or body part size or shape (such as, for example, breast augmentation or diminution, change in bone form or shape); effecting biorhythms or circadian cycles or rhythms; effecting the fertility of male or female subjects; effecting the metabolism, catabolism, anabolism, processing, utilization, storage or

elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, co-factors or other nutritional factors or component(s); effecting behavioral characteristics, including, without limitation, appetite, libido, stress, cognition (including cognitive disorders), depression (including depressive disorders) and violent behaviors; providing analgesic effects or other pain reducing effects; promoting differentiation and growth of embryonic stem cells in lineages other than hematopoietic lineages; hormonal or endocrine activity; in the case of enzymes, correcting deficiencies of the enzyme and treating deficiency-related diseases; treatment of hyperproliferative disorders (such as, for example, psoriasis); immunoglobulin-like activity (such as, for example, the ability to bind antigens or complement); and the ability to act as an antigen in a vaccine composition to raise an immune response against such protein or another material or entity which is cross-reactive with such protein.

4.10.19 IDENTIFICATION OF POLYMORPHISMS

The demonstration of polymorphisms makes possible the identification of such polymorphisms in human subjects and the pharmacogenetic use of this information for diagnosis and treatment. Such polymorphisms may be associated with, e.g., differential predisposition or susceptibility to various disease states (such as disorders involving inflammation or immune response) or a differential response to drug administration, and this genetic information can be used to tailor preventive or therapeutic treatment appropriately. For example, the existence of a polymorphism associated with a predisposition to inflammation or autoimmune disease makes possible the diagnosis of this condition in humans by identifying the presence of the polymorphism.

Polymorphisms can be identified in a variety of ways known in the art which all generally involve obtaining a sample from a patient, analyzing DNA from the sample, optionally involving isolation or amplification of the DNA, and identifying the presence of the polymorphism in the DNA. For example, PCR may be used to amplify an appropriate fragment of genomic DNA which may then be sequenced. Alternatively, the DNA may be subjected to allele-specific oligonucleotide hybridization (in which appropriate oligonucleotides are hybridized to the DNA under conditions permitting detection of a single base mismatch) or to a single nucleotide extension assay (in which an oligonucleotide that hybridizes immediately adjacent to the position of the polymorphism is extended with one or more labeled nucleotides). In addition, traditional restriction fragment length polymorphism analysis (using restriction enzymes that provide differential digestion of the genomic DNA depending on the presence or absence of the polymorphism) may be performed. Arrays with nucleotide sequences of the present invention can be used to detect polymorphisms. The array can comprise modified

nucleotide sequences of the present invention in order to detect the nucleotide sequences of the present invention. In the alternative, any one of the nucleotide sequences of the present invention can be placed on the array to detect changes from those sequences.

Alternatively a polymorphism resulting in a change in the amino acid sequence could also be detected by detecting a corresponding change in amino acid sequence of the protein, e.g., by an antibody specific to the variant sequence.

4.10.20 ARTHRITIS AND INFLAMMATION

The immunosuppressive effects of the compositions of the invention against rheumatoid arthritis is determined in an experimental animal model system. The experimental model system is adjuvant induced arthritis in rats, and the protocol is described by J. Holoshitz, et at., 1983, Science, 219:56, or by B. Waksman et al., 1963, Int. Arch. Allergy Appl. Immunol., 23:129. Induction of the disease can be caused by a single injection, generally intradermally, of a suspension of killed Mycobacterium tuberculosis in complete Freund's adjuvant (CFA). The route of injection can vary, but rats may be injected at the base of the tail with an adjuvant mixture. The polypeptide is administered in phosphate buffered solution (PBS) at a dose of about 1-5 mg/kg. The control consists of administering PBS only.

The procedure for testing the effects of the test compound would consist of intradermally injecting killed Mycobacterium tuberculosis in CFA followed by immediately administering the test compound and subsequent treatment every other day until day 24. At 14, 15, 18, 20, 22, and 24 days after injection of Mycobacterium CFA, an overall arthritis score may be obtained as described by J. Holoskitz above. An analysis of the data would reveal that the test compound would have a dramatic affect on the swelling of the joints as measured by a decrease of the arthritis score.

4.11 THERAPEUTIC METHODS

The compositions (including polypeptide fragments, analogs, variants and antibodies or other binding partners or modulators including antisense polynucleotides) of the invention have numerous applications in a variety of therapeutic methods. Examples of therapeutic applications include, but are not limited to, those exemplified herein.

4.11.1 EXAMPLE

One embodiment of the invention is the administration of an effective amount of the polypeptides or other composition of the invention to individuals affected by a disease or disorder that can be modulated by regulating the peptides of the invention. While the mode of

administration is not particularly important, parenteral administration is preferred. An exemplary mode of administration is to deliver an intravenous bolus. The dosage of the polypeptides or other composition of the invention will normally be determined by the prescribing physician. It is to be expected that the dosage will vary according to the age, weight, condition and response of the individual patient. Typically, the amount of polypeptide administered per dose will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight, with the preferred dose being about 0.1 μg/kg to 10 mg/kg of patient body weight. For parenteral administration, polypeptides of the invention will be formulated in an injectable form combined with a pharmaceutically acceptable parenteral vehicle. Such vehicles are well known in the art and examples include water, saline, Ringer's solution, dextrose solution, and solutions consisting of small amounts of the human serum albumin. The vehicle may contain minor amounts of additives that maintain the isotonicity and stability of the polypeptide or other active ingredient. The preparation of such solutions is within the skill of the art.

4.12 PHARMACEUTICAL FORMULATIONS AND ROUTES OF ADMINISTRATION

A protein or other composition of the present invention (from whatever source derived, including without limitation from recombinant and non-recombinant sources and including antibodies and other binding partners of the polypeptides of the invention) may be administered to a patient in need, by itself, or in pharmaceutical compositions where it is mixed with suitable carriers or excipient(s) at doses to treat or ameliorate a variety of disorders. Such a composition may optionally contain (in addition to protein or other active ingredient and a carrier) diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials well known in the art. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredient(s). The characteristics of the carrier will depend on the route of administration. The pharmaceutical composition of the invention may also contain cytokines, lymphokines, or other hematopoietic factors such as M-CSF, GM-CSF, TNF, IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IFN, TNF0, TNF1, TNF2, G-CSF, Meg-CSF, thrombopoietin, stem cell factor, and erythropoietin. In further compositions, proteins of the invention may be combined with other agents beneficial to the treatment of the disease or disorder in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet-derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), insulin-like growth factor (IGF), as well as cytokines described herein.

The pharmaceutical composition may further contain other agents which either enhance the activity of the protein or other active ingredient or complement its activity or use in treatment. Such additional factors and/or agents may be included in the pharmaceutical composition to produce a synergistic effect with protein or other active ingredient of the invention, or to minimize side effects. Conversely, protein or other active ingredient of the present invention may be included in formulations of the particular clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent to minimize side effects of the clotting factor, cytokine, lymphokine, other hematopoietic factor, thrombolytic or anti-thrombotic factor, or anti-inflammatory agent (such as IL-1Ra, IL-1 Hy1, IL-1 Hy2, anti-TNF, corticosteroids, immunosuppressive agents). A protein of the present invention may be active in multimers (e.g., heterodimers or homodimers) or complexes with itself or other proteins. As a result, pharmaceutical compositions of the invention may comprise a protein of the invention in such multimeric or complexed form.

As an alternative to being included in a pharmaceutical composition of the invention including a first protein, a second protein or a therapeutic agent may be concurrently administered with the first protein (e.g., at the same time, or at differing times provided that therapeutic concentrations of the combination of agents is achieved at the treatment site). Techniques for formulation and administration of the compounds of the instant application may be found in "Remington's Pharmaceutical Sciences," Mack Publishing Co., Easton, PA, latest edition. A therapeutically effective dose further refers to that amount of the compound sufficient to result in amelioration of symptoms, *e.g.*, treatment, healing, prevention or amelioration of the relevant medical condition, or an increase in rate of treatment, healing, prevention or amelioration of such conditions. When applied to an individual active ingredient, administered alone, a therapeutically effective dose refers to that ingredient alone. When applied to a combination, a therapeutically effective dose refers to combined amounts of the active ingredients that result in the therapeutic effect, whether administered in combination, serially or simultaneously.

In practicing the method of treatment or use of the present invention, a therapeutically effective amount of protein or other active ingredient of the present invention is administered to a mammal having a condition to be treated. Protein or other active ingredient of the present invention may be administered in accordance with the method of the invention either alone or in combination with other therapies such as treatments employing cytokines, lymphokines or other hematopoietic factors. When co- administered with one or more cytokines, lymphokines or other hematopoietic factors, protein or other active ingredient of the present invention may be administered either simultaneously with the cytokine(s), lymphokine(s), other hematopoietic

factor(s), thrombolytic or anti-thrombotic factors, or sequentially. If administered sequentially, the attending physician will decide on the appropriate sequence of administering protein or other active ingredient of the present invention in combination with cytokine(s), lymphokine(s), other hematopoietic factor(s), thrombolytic or anti-thrombotic factors.

4.12.1 ROUTES OF ADMINISTRATION

Suitable routes of administration may, for example, include oral, rectal, transmucosal, or intestinal administration; parenteral delivery, including intramuscular, subcutaneous, intramedullary injections, as well as intrathecal, direct intraventricular, intravenous, intraperitoneal, intranasal, or intraocular injections. Administration of protein or other active ingredient of the present invention used in the pharmaceutical composition or to practice the method of the present invention can be carried out in a variety of conventional ways, such as oral ingestion, inhalation, topical application or cutaneous, subcutaneous, intraperitoneal, parenteral or intravenous injection. Intravenous administration to the patient is preferred.

Alternately, one may administer the compound in a local rather than systemic manner, for example, via injection of the compound directly into a arthritic joints or in fibrotic tissue, often in a depot or sustained release formulation. In order to prevent the scarring process frequently occurring as complication of glaucoma surgery, the compounds may be administered topically, for example, as eye drops. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a specific antibody, targeting, for example, arthritic or fibrotic tissue. The liposomes will be targeted to and taken up selectively by the afflicted tissue.

The polypeptides of the invention are administered by any route that delivers an effective dosage to the desired site of action. The determination of a suitable route of administration and an effective dosage for a particular indication is within the level of skill in the art. Preferably for wound treatment, one administers the therapeutic compound directly to the site. Suitable dosage ranges for the polypeptides of the invention can be extrapolated from these dosages or from similar studies in appropriate animal models. Dosages can then be adjusted as necessary by the clinician to provide maximal therapeutic benefit.

4.12.2 COMPOSITIONS/FORMULATIONS

Pharmaceutical compositions for use in accordance with the present invention thus may be formulated in a conventional manner using one or more physiologically acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. These pharmaceutical compositions may be

manufactured in a manner that is itself known, *e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes. Proper formulation is dependent upon the route of administration chosen. When a therapeutically effective amount of protein or other active ingredient of the present invention is administered orally, protein or other active ingredient of the present invention will be in the form of a tablet, capsule, powder, solution or elixir. When administered in tablet form, the pharmaceutical composition of the invention may additionally contain a solid carrier such as a gelatin or an adjuvant. The tablet, capsule, and powder contain from about 5 to 95% protein or other active ingredient of the present invention, and preferably from about 25 to 90% protein or other active ingredient of the present invention. When administered in liquid form, a liquid carrier such as water, petroleum, oils of animal or plant origin such as peanut oil, mineral oil, soybean oil, or sesame oil, or synthetic oils may be added. The liquid form of the pharmaceutical composition may further contain physiological saline solution, dextrose or other saccharide solution, or glycols such as ethylene glycol, propylene glycol or polyethylene glycol. When administered in liquid form, the pharmaceutical composition contains from about 0.5 to 90% by weight of protein or other active ingredient of the present invention, and preferably from about 1 to 50% protein or other active ingredient of the present invention.

When a therapeutically effective amount of protein or other active ingredient of the present invention is administered by intravenous, cutaneous or subcutaneous injection, protein or other active ingredient of the present invention will be in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such parenterally acceptable protein or other active ingredient solutions, having due regard to pH, isotonicity, stability, and the like, is within the skill in the art. A preferred pharmaceutical composition for intravenous, cutaneous, or subcutaneous injection should contain, in addition to protein or other active ingredient of the present invention, an isotonic vehicle such as Sodium Chloride Injection, Ringer's Injection, Dextrose Injection, Dextrose and Sodium Chloride Injection, Lactated Ringer's Injection, or other vehicle as known in the art. The pharmaceutical composition of the present invention may also contain stabilizers, preservatives, buffers, antioxidants, or other additives known to those of skill in the art. For injection, the agents of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiological saline buffer. For transmucosal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

For oral administration, the compounds can be formulated readily by combining the active compounds with pharmaceutically acceptable carriers well known in the art. Such carriers

enable the compounds of the invention to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a patient to be treated. Pharmaceutical preparations for oral use can be obtained from a solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch, gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Dragee cores are provided with suitable coatings. For this purpose, concentrated sugar solutions may be used, which may optionally contain gum arabic, talc, polyvinyl pyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for identification or to characterize different combinations of active compound doses.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. All formulations for oral administration should be in dosages suitable for such administration. For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

For administration by inhalation, the compounds for use according to the present invention are conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebuliser, with the use of a suitable propellant, *e.g.*, dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined by providing a valve to deliver a metered amount. Capsules and cartridges of, *e.g.*, gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch. The compounds may be formulated for parenteral administration by injection, *e.g.*, by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, *e.g.*, in ampules or in multi-dose containers, with

an added preservative. The compositions may take such forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. Alternatively, the active ingredient may be in powder form for constitution with a suitable vehicle, *e.g.*, sterile pyrogen-free water, before use.

The compounds may also be formulated in rectal compositions such as suppositories or retention enemas, *e.g.*, containing conventional suppository bases such as cocoa butter or other glycerides. In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be administered by implantation (for example subcutaneously or intramuscularly) or by intramuscular injection. Thus, for example, the compounds may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

A pharmaceutical carrier for the hydrophobic compounds of the invention is a co-solvent system comprising benzyl alcohol, a nonpolar surfactant, a water-miscible organic polymer, and an aqueous phase. The co-solvent system may be the VPD co-solvent system. VPD is a solution of 3% w/v benzyl alcohol, 8% w/v of the nonpolar surfactant polysorbate 80, and 65% w/v polyethylene glycol 300, made up to volume in absolute ethanol. The VPD co-solvent system (VPD:5W) consists of VPD diluted 1:1 with a 5% dextrose in water solution. This co-solvent system dissolves hydrophobic compounds well, and itself produces low toxicity upon systemic administration. Naturally, the proportions of a co-solvent system may be varied considerably without destroying its solubility and toxicity characteristics. Furthermore, the identity of the co-solvent components may be varied: for example, other low-toxicity nonpolar surfactants may be used instead of polysorbate 80; the fraction size of polyethylene glycol may be varied; other biocompatible polymers may replace polyethylene glycol, e.g. polyvinyl pyrrolidone; and other sugars or polysaccharides may substitute for dextrose. Alternatively, other delivery systems for hydrophobic pharmaceutical compounds may be employed. Liposomes and emulsions are well

known examples of delivery vehicles or carriers for hydrophobic drugs. Certain organic solvents such as dimethylsulfoxide also may be employed, although usually at the cost of greater toxicity. Additionally, the compounds may be delivered using a sustained-release system, such as semipermeable matrices of solid hydrophobic polymers containing the therapeutic agent. Various types of sustained-release materials have been established and are well known by those skilled in the art. Sustained-release capsules may, depending on their chemical nature, release the compounds for a few weeks up to over 100 days. Depending on the chemical nature and the biological stability of the therapeutic reagent, additional strategies for protein or other active ingredient stabilization may be employed.

The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols. Many of the active ingredients of the invention may be provided as salts with pharmaceutically compatible counter ions. Such pharmaceutically acceptable base addition salts are those salts which retain the biological effectiveness and properties of the free acids and which are obtained by reaction with inorganic or organic bases such as sodium hydroxide, magnesium hydroxide, ammonia, trialkylamine, dialkylamine, monoalkylamine, dibasic amino acids, sodium acetate, potassium benzoate, triethanol amine and the like.

The pharmaceutical composition of the invention may be in the form of a complex of the protein(s) or other active ingredient(s) of present invention along with protein or peptide antigens. The protein and/or peptide antigen will deliver a stimulatory signal to both B and T lymphocytes. B lymphocytes will respond to antigen through their surface immunoglobulin receptor. T lymphocytes will respond to antigen through the T cell receptor (TCR) following presentation of the antigen by MHC proteins. MHC and structurally related proteins including those encoded by class I and class II MHC genes on host cells will serve to present the peptide antigen(s) to T lymphocytes. The antigen components could also be supplied as purified MHC-peptide complexes alone or with co-stimulatory molecules that can directly signal T cells. Alternatively antibodies able to bind surface immunoglobulin and other molecules on B cells as well as antibodies able to bind the TCR and other molecules on T cells can be combined with the pharmaceutical composition of the invention.

The pharmaceutical composition of the invention may be in the form of a liposome in which protein of the present invention is combined, in addition to other pharmaceutically acceptable carriers, with amphipathic agents such as lipids which exist in aggregated form as micelles, insoluble monolayers, liquid crystals, or lamellar layers in aqueous solution. Suitable

lipids for liposomal formulation include, without limitation, monoglycerides, diglycerides, sulfatides, lysolecithins, phospholipids, saponin, bile acids, and the like. Preparation of such liposomal formulations is within the level of skill in the art, as disclosed, for example, in U.S. Patent Nos. 4,235,871; 4,501,728; 4,837,028; and 4,737,323, all of which are incorporated herein by reference.

The amount of protein or other active ingredient of the present invention in the pharmaceutical composition of the present invention will depend upon the nature and severity of the condition being treated, and on the nature of prior treatments which the patient has undergone. Ultimately, the attending physician will decide the amount of protein or other active ingredient of the present invention with which to treat each individual patient. Initially, the attending physician will administer low doses of protein or other active ingredient of the present invention and observe the patient's response. Larger doses of protein or other active ingredient of the present invention may be administered until the optimal therapeutic effect is obtained for the patient, and at that point the dosage is not increased further. It is contemplated that the various pharmaceutical compositions used to practice the method of the present invention should contain about 0.01 μg to about 100 mg (preferably about 0.1 μg to about 10 mg, more preferably about 0.1 μg to about 1 mg) of protein or other active ingredient of the present invention per kg body weight. For compositions of the present invention which are useful for bone, cartilage, tendon or ligament regeneration, the therapeutic method includes administering the composition topically, systematically, or locally as an implant or device. When administered, the therapeutic composition for usc in this invention is, of course, in a pyrogen-free, physiologically acceptable form. Further, the composition may desirably be encapsulated or injected in a viscous form for delivery to the site of bone, cartilage or tissue damage. Topical administration may be suitable for wound healing and tissue repair. Therapeutically useful agents other than a protein or other active ingredient of the invention which may also optionally be included in the composition as described above, may alternatively or additionally, be administered simultaneously or sequentially with the composition in the methods of the invention. Preferably for bone and/or cartilage formation, the composition would include a matrix capable of delivering the protein-containing or other active ingredient-containing composition to the site of bone and/or cartilage damage, providing a structure for the developing bone and cartilage and optimally capable of being resorbed into the body. Such matrices may be formed of materials presently in use for other implanted medical applications.

The choice of matrix material is based on biocompatibility, biodegradability, mechanical properties, cosmetic appearance and interface properties. The particular application of the compositions will define the appropriate formulation. Potential matrices for the compositions

may be biodegradable and chemically defined calcium sulfate, tricalcium phosphate, hydroxyapatite, polylactic acid, polyglycolic acid and polyanhydrides. Other potential materials are biodegradable and biologically well-defined, such as bone or dermal collagen. Further matrices are comprised of pure proteins or extracellular matrix components. Other potential matrices are nonbiodegradable and chemically defined, such as sintered hydroxyapatite, bioglass, aluminates, or other ceramics. Matrices may be comprised of combinations of any of the above mentioned types of material, such as polylactic acid and hydroxyapatite or collagen and tricalcium phosphate. The bioceramics may be altered in composition, such as in calcium-aluminate-phosphate and processing to alter pore size, particle size, particle shape, and biodegradability. Presently preferred is a 50:50 (mole weight) copolymer of lactic acid and glycolic acid in the form of porous particles having diameters ranging from 150 to 800 microns. In some applications, it will be useful to utilize a sequestering agent, such as carboxymethyl cellulose or autologous blood clot, to prevent the protein compositions from disassociating from the matrix.

A preferred family of sequestering agents is cellulosic materials such as alkylcelluloses (including hydroxyalkylcelluloses), including methylcellulose, ethylcellulose, hydroxyethylcellulose, hydroxypropylcellulose, hydroxypropyl-methylcellulose, and carboxymethylcellulose, the most preferred being cationic salts of carboxymethylcellulose (CMC). Other preferred sequestering agents include hyaluronic acid, sodium alginate, poly(ethylene glycol), polyoxyethylene oxide, carboxyvinyl polymer and poly(vinyl alcohol). The amount of sequestering agent useful herein is 0.5-20 wt %, preferably 1-10 wt % based on total formulation weight, which represents the amount necessary to prevent desorption of the protein from the polymer matrix and to provide appropriate handling of the composition, yet not so much that the progenitor cells are prevented from infiltrating the matrix, thereby providing the protein the opportunity to assist the osteogenic activity of the progenitor cells. In further compositions, proteins or other active ingredients of the invention may be combined with other agents beneficial to the treatment of the bone and/or cartilage defect, wound, or tissue in question. These agents include various growth factors such as epidermal growth factor (EGF), platelet derived growth factor (PDGF), transforming growth factors (TGF-α and TGF-β), and insulin-like growth factor (IGF).

The therapeutic compositions are also presently valuable for veterinary applications. Particularly domestic animals and thoroughbred horses, in addition to humans, are desired patients for such treatment with proteins or other active ingredients of the present invention. The dosage regimen of a protein-containing pharmaceutical composition to be used in tissue regeneration will be determined by the attending physician considering various factors which

modify the action of the proteins, *e.g.*, amount of tissue weight desired to be formed, the site of damage, the condition of the damaged tissue, the size of a wound, type of damaged tissue (*e.g.*, bone), the patient's age, sex, and diet, the severity of any infection, time of administration and other clinical factors. The dosage may vary with the type of matrix used in the reconstitution and with inclusion of other proteins in the pharmaceutical composition. For example, the addition of other known growth factors, such as IGF I (insulin like growth factor I), to the final composition, may also effect the dosage. Progress can be monitored by periodic assessment of tissue/bone growth and/or repair, for example, X-rays, histomorphometric determinations and tetracycline labeling.

Polynucleotides of the present invention can also be used for gene therapy. Such polynucleotides can be introduced either in vivo or ex vivo into cells for expression in a mammalian subject. Polynucleotides of the invention may also be administered by other known methods for introduction of nucleic acid into a cell or organism (including, without limitation, in the form of viral vectors or naked DNA). Cells may also be cultured ex vivo in the presence of proteins of the present invention in order to proliferate or to produce a desired effect on or activity in such cells. Treated cells can then be introduced in vivo for therapeutic purposes.

4.12.3 EFFECTIVE DOSAGE

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve its intended purpose. More specifically, a therapeutically effective amount means an amount effective to prevent development of or to alleviate the existing symptoms of the subject being treated. Determination of the effective amount is well within the capability of those skilled in the art, especially in light of the detailed disclosure provided herein. For any compound used in the method of the invention, the therapeutically effective dose can be estimated initially from appropriate in vitro assays. For example, a dose can be formulated in animal models to achieve a circulating concentration range that can be used to more accurately determine useful doses in humans. For example, a dose can be formulated in animal models to achieve a circulating concentration range that includes the IC_{50} as determined in cell culture (*i.e.*, the concentration of the test compound which achieves a half-maximal inhibition of the protein's biological activity). Such information can be used to more accurately determine useful doses in humans.

A therapeutically effective dose refers to that amount of the compound that results in amelioration of symptoms or a prolongation of survival in a patient. Toxicity and therapeutic efficacy of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, *e.g.*, for determining the LD_{50} (the dose lethal to 50% of the

population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio between LD_{50} and ED_{50}. Compounds which exhibit high therapeutic indices are preferred. The data obtained from these cell culture assays and animal studies can be used in formulating a range of dosage for use in human. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. The exact formulation, route of administration and dosage can be chosen by the individual physician in view of the patient's condition. See, *e.g.*, Fingl et al., 1975, in "The Pharmacological Basis of Therapeutics", Ch. 1 p.1. Dosage amount and interval may be adjusted individually to provide plasma levels of the active moiety which are sufficient to maintain the desired effects, or minimal effective concentration (MEC). The MEC will vary for each compound but can be estimated from *in vitro* data. Dosages necessary to achieve the MEC will depend on individual characteristics and route of administration. However, HPLC assays or bioassays can be used to determine plasma concentrations.

Dosage intervals can also be determined using MEC value. Compounds should be administered using a regimen which maintains plasma levels above the MEC for 10-90% of the time, preferably between 30-90% and most preferably between 50-90%. In cases of local administration or selective uptake, the effective local concentration of the drug may not be related to plasma concentration.

An exemplary dosage regimen for polypeptides or other compositions of the invention will be in the range of about 0.01 μg/kg to 100 mg/kg of body weight daily, with the preferred dose being about 0.1 μg/kg to 25 mg/kg of patient body weight daily, varying in adults and children. Dosing may be once daily, or equivalent doses may be delivered at longer or shorter intervals.

The amount of composition administered will, of course, be dependent on the subject being treated, on the subject's age and weight, the severity of the affliction, the manner of administration and the judgment of the prescribing physician.

4.12.4 PACKAGING

The compositions may, if desired, be presented in a pack or dispenser device which may contain one or more unit dosage forms containing the active ingredient. The pack may, for example, comprise metal or plastic foil, such as a blister pack. The pack or dispenser device may be accompanied by instructions for administration. Compositions comprising a compound of the

invention formulated in a compatible pharmaceutical carrier may also be prepared, placed in an appropriate container, and labeled for treatment of an indicated condition.

4.13 ANTIBODIES

Also included in the invention are antibodies to proteins, or fragments of proteins of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, i.e., molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab}, $F_{ab'}$ and $F_{(ab')2}$ fragments, and an F_{ab} expression library. In general, an antibody molecule obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG_1, IgG_2, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated related protein of the invention may be intended to serve as an antigen, or a portion or fragment thereof, and additionally can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence shown in SEQ ID NO: 173, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of -related protein that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human related protein sequence will indicate which regions of a related protein are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte

Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each of which is incorporated herein by reference in its entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, Antibodies: A Laboratory Manual, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

4.13.1 POLYCLONAL ANTIBODIES

For the production of polyclonal antibodies, various suitable host animals (e.g., rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (e.g., aluminum hydroxide), surface active substances (e.g., lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, etc.), adjuvants usable in humans such as Bacille Calmette-Guerin and Corynebacterium parvum, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (e.g., from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the

target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

4.13.2 MONOCLONAL ANTIBODIES

The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MAbs thus contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro. The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center, San Diego,

California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). Preferably, antibodies having a high degree of specificity and a high binding affinity for the target antigen are isolated.

After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.
The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatitc chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin

polypeptide can be substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

4.13.3 HUMANIZED ANTIBODIES

The antibodies directed against the protein antigens of the invention can further comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', $F(ab')_2$ or other antigen-binding subsequences of antibodies) that are principally comprised of the sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeyen et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

4.13.4 HUMAN ANTIBODIES

Fully human antibodies relate to antibody molecules in which essentially the entire sequences of both the light chain and the heavy chain, including the CDRs, arise from human genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein. Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal

antibodies may be utilized in the practice of the present invention and may be produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96).

In addition, human antibodies can also be produced using additional techniques, including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991); Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by introducing human immunoglobulin loci into transgenic animals, e.g., mice in which the endogenous immunoglobulin genes have been partially or completely inactivated. Upon challenge, human antibody production is observed, which closely resembles that seen in humans in all respects, including gene rearrangement, assembly, and antibody repertoire. This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806; 5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10, 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368, 812-13 (1994)); Fishwild et al,(Nature Biotechnology 14, 845-51 (1996)); Neuberger (Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13 65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals which are modified so as to produce fully human antibodies rather than the animal's endogenous antibodies in response to challenge by an antigen. (See PCT publication WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the Xenomouse™ as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

4.13.5 F_{ab} FRAGMENTS AND SINGLE CHAIN ANTIBODIES

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see e.g., Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotypes to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an $F_{(ab')2}$ fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an $F_{(ab')2}$ fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

4.13.6 BISPECIFIC ANTIBODIES

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the

binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker *et al.*, 1991 *EMBO J.*, 10:3655-3659.

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., Methods in Enzymology, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (e.g. tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (e.g. alanine or threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (e.g. $F(ab')_2$ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies can be prepared using chemical linkage. Brennan et al., Science 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate $F(ab')_2$ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to

stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from E. coli and chemically coupled to form bispecific antibodies. Shalaby et al., J. Exp. Med. 175:217-225 (1992) describe the production of a fully humanized bispecific antibody $F(ab')_2$ molecule. Each Fab' fragment was separately secreted from E. coli and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers. The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (e.g. CD2, CD3, CD28, or B7), or Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular

defense mechanisms to the cell expressing the particular antigen. Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA. Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

4.13.7 HETEROCONJUGATE ANTIBODIES

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

4.13.8 EFFECTOR FUNCTION ENGINEERING

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, e.g., the effectiveness of the antibody in treating cancer. For example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

4.13.9 IMMUNOCONJUGATES

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (e.g., an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (i.e., a radioconjugate).

a nucleotide sequence of the present invention. As used herein, "recorded" refers to a process for storing information on computer readable medium. A skilled artisan can readily adopt any of the presently known methods for recording information on computer readable medium to generate manufactures comprising the nucleotide sequence information of the present invention.

A variety of data storage structures are available to a skilled artisan for creating a computer readable medium having recorded thereon a nucleotide sequence of the present invention. The choice of the data storage structure will generally be based on the means chosen to access the stored information. In addition, a variety of data processor programs and formats can be used to store the nucleotide sequence information of the present invention on computer readable medium. The sequence information can be represented in a word processing text file, formatted in commercially-available software such as WordPerfect and Microsoft Word, or represented in the form of an ASCII file, stored in a database application, such as DB2, Sybase, Oracle, or the like. A skilled artisan can readily adapt any number of data processor structuring formats (*e.g.* text file or database) in order to obtain computer readable medium having recorded thereon the nucleotide sequence information of the present invention.

By providing any of the nucleotide sequences SEQ ID NO: 1-172, or 345-516 or a representative fragment thereof; or a nucleotide sequence at least 95% identical to any of the nucleotide sequences of SEQ ID NO: 1-172, or 345-516 in computer readable form, a skilled artisan can routinely access the sequence information for a variety of purposes. Computer software is publicly available which allows a skilled artisan to access sequence information provided in a computer readable medium. The examples which follow demonstrate how software which implements the BLAST (Altschul et al., J. Mol. Biol. 215:403-410 (1990)) and BLAZE (Brutlag et al., Comp. Chem. 17:203-207 (1993)) search algorithms on a Sybase system is used to identify open reading frames (ORFs) within a nucleic acid sequence. Such ORFs may be protein encoding fragments and may be useful in producing commercially important proteins such as enzymes used in fermentation reactions and in the production of commercially useful metabolites.

As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware means of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. As stated above, the computer-based systems of the present invention comprise a data storage means having stored therein a nucleotide sequence of the present invention and the necessary hardware means and

software means for supporting and implementing a search means. As used herein, "data storage means" refers to memory which can store nucleotide sequence information of the present invention, or a memory access means which can access manufactures having recorded thereon the nucleotide sequence information of the present invention.

As used herein, "search means" refers to one or more programs which are implemented on the computer-based system to compare a target sequence or target structural motif with the sequence information stored within the data storage means. Search means are used to identify fragments or regions of a known sequence which match a particular target sequence or target motif. A variety of known algorithms are disclosed publicly and a variety of commercially available software for conducting search means are and can be used in the computer-based systems of the present invention. Examples of such software includes, but is not limited to, Smith-Waterman, MacPattern (EMBL), BLASTN and BLASTA (NPOLYPEPTIDEIA). A skilled artisan can readily recognize that any one of the available algorithms or implementing software packages for conducting homology searches can be adapted for use in the present computer-based systems. As used herein, a "target sequence" can be any nucleic acid or amino acid sequence of six or more nucleotides or two or more amino acids. A skilled artisan can readily recognize that the longer a target sequence is, the less likely a target sequence will be present as a random occurrence in the database. The most preferred sequence length of a target sequence is from about 10 to 300 amino acids, more preferably from about 30 to 100 nucleotide residues. However, it is well recognized that searches for commercially important fragments, such as sequence fragments involved in gene expression and protein processing, may be of shorter length.

As used herein, "a target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration which is formed upon the folding of the target motif. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, promoter sequences, hairpin structures and inducible expression elements (protein binding sequences).

4.15 TRIPLE HELIX FORMATION

In addition, the fragments of the present invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on the binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for usc in these methods are preferably 20 to 40 bases in length and are

designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 15241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Olmno, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

4.16 DIAGNOSTIC ASSAYS AND KITS

The present invention further provides methods to identify the presence or expression of one of the ORFs of the present invention, or homolog thereof, in a test sample, using a nucleic acid probe or antibodies of the present invention, optionally conjugated or otherwise associated with a suitable label.

In general, methods for detecting a polynucleotide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polynucleotide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polynucleotide of the invention is detected in the sample. Such methods can also comprise contacting a sample under stringent hybridization conditions with nucleic acid primers that anneal to a polynucleotide of the invention under such conditions, and amplifying annealed polynucleotides, so that if a polynucleotide is amplified, a polynucleotide of the invention is detected in the sample.

In general, methods for detecting a polypeptide of the invention can comprise contacting a sample with a compound that binds to and forms a complex with the polypeptide for a period sufficient to form the complex, and detecting the complex, so that if a complex is detected, a polypeptide of the invention is detected in the sample.

In detail, such methods comprise incubating a test sample with one or more of the antibodies or one or more of the nucleic acid probes of the present invention and assaying for binding of the nucleic acid probes or antibodies to components within the test sample.

Conditions for incubating a nucleic acid probe or antibody with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid probe or antibody used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or immunological assay formats can readily be adapted to employ the nucleic acid

probes or antibodies of the present invention. Examples of such assays can be found in Chard, T., An Introduction to Radioimmunoassay and Related Techniques, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G.R. et al., Tcchniques in Immunocytochemistry, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., Practice and Theory of immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology, Elsevier Science Publishers, Amsterdam, The Netherlands (1985). The test samples of the present invention include cells, protein or membrane extracts of cells, or biological fluids such as sputum, blood, serum, plasma, or urine. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing protein extracts or membrane extracts of cells are well known in the art and can be readily be adapted in order to obtain a sample which is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention. Specifically, the invention provides a compartment kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the probes or antibodies of the present invention; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound probe or antibody.

In detail, a compartment kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers or strips of plastic or paper. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the antibodies used in the assay, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound antibody or probe. Types of detection reagents include labeled nucleic acid probes, labeled secondary antibodies, or in the alternative, if the primary antibody is labeled, the enzymatic, or antibody binding reagents which are capable of reacting with the labeled antibody. One skilled in the art will readily recognize that the disclosed probes and antibodies of the present invention can be readily incorporated into one of the established kit formats which are well known in the art.

4.17 MEDICAL IMAGING

The novel polypeptides and binding partners of the invention are useful in medical imaging of sites expressing the molecules of the invention (e.g., where the polypeptide of the invention is involved in the immune response, for imaging sites of inflammation or infection). See, e.g., Kunkel et al., U.S. Pat. NO. 5,413,778. Such methods involve chemical attachment of a labeling or imaging agent, administration of the labeled polypeptide to a subject in a pharmaceutically acceptable carrier, and imaging the labeled polypeptide *in vivo* at the target site.

4.18 SCREENING ASSAYS

Using the isolated proteins and polynucleotides of the invention, the present invention further provides methods of obtaining and identifying agents which bind to a polypeptide encoded by an ORF corresponding to any of the nucleotide sequences set forth in SEQ ID NO: 1-172, or 345-516, or bind to a specific domain of the polypeptide encoded by the nucleic acid. In detail, said method comprises the steps of:

(a) contacting an agent with an isolated protein encoded by an ORF of the present invention, or nucleic acid of the invention; and

(b) determining whether the agent binds to said protein or said nucleic acid.

In general, therefore, such methods for identifying compounds that bind to a polynucleotide of the invention can comprise contacting a compound with a polynucleotide of the invention for a time sufficient to form a polynucleotide/compound complex, and detecting the complex, so that if a polynucleotide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified.

Likewise, in general, therefore, such methods for identifying compounds that bind to a polypeptide of the invention can comprise contacting a compound with a polypeptide of the invention for a time sufficient to form a polypeptide/compound complex, and detecting the complex, so that if a polypeptide/compound complex is detected, a compound that binds to a polynucleotide of the invention is identified..

Methods for identifying compounds that bind to a polypeptide of the invention can also comprise contacting a compound with a polypeptide of the invention in a cell for a time sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a receptor gene sequence in the cell, and detecting the complex by detecting reporter gene sequence expression, so that if a polypeptide/compound complex is detected, a compound that binds a polypeptide of the invention is identified.

Compounds identified via such methods can include compounds which modulate the activity of a polypeptide of the invention (that is, increase or decrease its activity, relative to

activity observed in the absence of the compound). Alternatively, compounds identified via such methods can include compounds which modulate the expression of a polynucleotide of the invention (that is, increase or decrease expression relative to expression levels observed in the absence of the compound). Compounds, such as compounds identified via the methods of the invention, can be tested using standard assays well known to those of skill in the art for their ability to modulate activity/expression.

The agents screened in the above assay can be, but are not limited to, peptides, carbohydrates, vitamin derivatives, or other pharmaceutical agents. The agents can be selected and screened at random or rationally selected or designed using protein modeling techniques.

For random screening, agents such as peptides, carbohydrates, pharmaceutical agents and the like are selected at random and are assayed for their ability to bind to the protein encoded by the ORF of the present invention. Alternatively, agents may be rationally selected or designed. As used herein, an agent is said to be "rationally selected or designed" when the agent is chosen based on the configuration of the particular protein. For example, one skilled in the art can readily adapt currently available procedures to generate peptides, pharmaceutical agents and the like, capable of binding to a specific peptide sequence, in order to generate rationally designed antipeptide peptides, for example see Hurby et al., Application of Synthetic Peptides: Antisense Peptides," In Synthetic Peptides, A User's Guide, W.H. Freeman, NY (1992), pp. 289-307, and Kaspczak et al., Biochemistry 28:9230-8 (1989), or pharmaceutical agents, or the like.

In addition to the foregoing, one class of agents of the present invention, as broadly described, can be used to control gene expression through binding to one of the ORFs or EMFs of the present invention. As described above, such agents can be randomly screened or rationally designed/selected. Targeting the ORF or EMF allows a skilled artisan to design sequence specific or element specific agents, modulating the expression of either a single ORF or multiple ORFs which rely on the same EMF for expression control. One class of DNA binding agents are agents which contain base residues which hybridize or form a triple helix formation by binding to DNA or RNA. Such agents can be based on the classic phosphodiester, ribonucleic acid backbone, or can be a variety of sulfhydryl or polymeric derivatives which have base attachment capacity.

Agents suitable for use in these methods preferably contain 20 to 40 bases and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). Triple helix-formation optimally results in a shut-off of RNA transcription

from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be effective in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide and other DNA binding agents.

Agents which bind to a protein encoded by one of the ORFs of the present invention can be used as a diagnostic agent. Agents which bind to a protein encoded by one of the ORFs of the present invention can be formulated using known techniques to generate a pharmaceutical composition.

4.19 USE OF NUCLEIC ACIDS AS PROBES

Another aspect of the subject invention is to provide for polypeptide-specific nucleic acid hybridization probes capable of hybridizing with naturally occurring nucleotide sequences. The hybridization probes of the subject invention may be derived from any of the nucleotide sequences SEQ ID NO: 1-172, or 345-516. Because the corresponding gene is only expressed in a limited number of tissues, a hybridization probe derived from of any of the nucleotide sequences SEQ ID NO: 1-172, or 345-516 can be used as an indicator of the presence of RNA of cell type of such a tissue in a sample.

Any suitable hybridization technique can be employed, such as, for example, in situ hybridization. PCR as described in US Patents Nos. 4,683,195 and 4,965,188 provides additional uses for oligonucleotides based upon the nucleotide sequences. Such probes used in PCR may be of recombinant origin, may be chemically synthesized, or a mixture of both. The probe will comprise a discrete nucleotide sequence for the detection of identical sequences or a degenerate pool of possible sequences for identification of closely related genomic sequences.

Other means for producing specific hybridization probes for nucleic acids include the cloning of nucleic acid sequences into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides. The nucleotide sequences may be used to construct hybridization probes for mapping their respective genomic sequences. The nucleotide sequence provided herein may be mapped to a chromosome or specific regions of a chromosome using well known genetic and/or chromosomal mapping techniques. These techniques include in situ hybridization, linkage analysis against known chromosomal markers, hybridization screening with libraries or flow-sorted chromosomal preparations specific to known chromosomes, and the like. The technique of fluorescent in situ hybridization of

chromosome spreads has been described, among other places, in Verma et al (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York NY.

Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques may be correlated with additional genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of a nucleic acid on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

4.20 PREPARATION OF SUPPORT BOUND OLIGONUCLEOTIDES

Oligonucleotides, i.e., small nucleic acid segments, may be readily prepared by, for example, directly synthesizing the oligonucleotide by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer.

Support bound oligonucleotides may be prepared by any of the methods known to those of skill in the art using any suitable support such as glass, polystyrene or Teflon. One strategy is to precisely spot oligonucleotides synthesized by standard synthesizers. Immobilization can be achieved using passive adsorption (Inouye & Hondo, (1990) J. Clin. Microbiol. 28(6) 1469-72); using UV light (Nagata *et al.*, 1985; Dahlen *et al.*, 1987; Morrissey & Collins, (1989) Mol. Cell Probes 3(2) 189-207) or by covalent binding of base modified DNA (Keller *et al.*, 1988; 1989); all references being specifically incorporated herein.

Another strategy that may be employed is the use of the strong biotin-streptavidin interaction as a linker. For example, Broude *et al.* (1994) Proc. Natl. Acad. Sci. USA 91(8) 3072-6, describe the use of biotinylated probes, although these are duplex probes, that are immobilized on streptavidin-coated magnetic beads. Streptavidin-coated beads may be purchased from Dynal, Oslo. Of course, this same linking chemistry is applicable to coating any surface with streptavidin. Biotinylated probes may be purchased from various sources, such as, e.g., Operon Technologies (Alameda, CA).

Nunc Laboratories (Naperville, IL) is also selling suitable material that could be used. Nunc Laboratories have developed a method by which DNA can be covalently bound to the microwell surface termed Covalink NH. CovaLink NH is a polystyrene surface grafted with secondary amino groups (>NH) that serve as bridge-heads for further covalent coupling. CovaLink Modules may be purchased from Nunc Laboratories. DNA molecules may be bound to CovaLink exclusively at the 5'-end by a phosphoramidate bond, allowing immobilization of more than 1 pmol of DNA (Rasmussen *et al.*, (1991) Anal. Biochem. 198(1) 138-42).

The use of CovaLink NH strips for covalent binding of DNA molecules at the 5'-end has been described (Rasmussen et al., (1991). In this technology, a phosphoramidate bond is employed (Chu et al., (1983) Nucleic Acids Res. 11(8) 6513-29). This is beneficial as immobilization using only a single covalent bond is preferred. The phosphoramidate bond joins the DNA to the CovaLink NH secondary amino groups that are positioned at the end of spacer arms covalently grafted onto the polystyrene surface through a 2 nm long spacer arm. To link an oligonucleotide to CovaLink NH via an phosphoramidate bond, the oligonucleotide terminus must have a 5'-end phosphate group. It is, perhaps, even possible for biotin to be covalently bound to CovaLink and then streptavidin used to bind the probes.

More specifically, the linkage method includes dissolving DNA in water (7.5 ng/ul) and denaturing for 10 min. at 95°C and cooling on ice for 10 min. Ice-cold 0.1 M 1-methylimidazole, pH 7.0 (1-$MeIm_7$), is then added to a final concentration of 10 mM 1-$MeIm_7$. A ss DNA solution is then dispensed into CovaLink NH strips (75 ul/well) standing on ice.

Carbodiimide 0.2 M 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (EDC), dissolved in 10 mM 1-$MeIm_7$, is made fresh and 25 ul added per well. The strips are incubated for 5 hours at 50°C. After incubation the strips are washed using, e.g., Nunc-Immuno Wash; first the wells are washed 3 times, then they are soaked with washing solution for 5 min., and finally they are washed 3 times (where in the washing solution is 0.4 N NaOH, 0.25% SDS heated to 50°C).

It is contemplated that a further suitable method for use with the present invention is that described in PCT Patent Application WO 90/03382 (Southern & Maskos), incorporated herein by reference. This method of preparing an oligonucleotide bound to a support involves attaching a nucleoside 3'-reagent through the phosphate group by a covalent phosphodiester link to aliphatic hydroxyl groups carried by the support. The oligonucleotide is then synthesized on the supported nucleoside and protecting groups removed from the synthetic oligonucleotide chain under standard conditions that do not cleave the oligonucleotide from the support. Suitable reagents include nucleoside phosphoramidite and nucleoside hydrogen phosphorate.

An on-chip strategy for the preparation of DNA probe for the preparation of DNA probe arrays may be employed. For example, addressable laser-activated photodeprotection may be employed in the chemical synthesis of oligonucleotides directly on a glass surface, as described by Fodor *et al.* (1991) Science 251(4995) 767-73, incorporated herein by reference. Probes may also be immobilized on nylon supports as described by Van Ness *et al.* (1991) Nucleic Acids Res. 19(12) 3345-50; or linked to Teflon using the method of Duncan & Cavalier (1988) Anal. Biochem. 169(1) 104-8; all references being specifically incorporated herein.

To link an oligonucleotide to a nylon support, as described by Van Ness *et al.* (1991), requires activation of the nylon surface via alkylation and selective activation of the 5'-amine of oligonucleotides with cyanuric chloride.

One particular way to prepare support bound oligonucleotides is to utilize the light-generated synthesis described by Pease *et al.*, (1994) PNAS USA 91(11) 5022-6, incorporated herein by reference). These authors used current photolithographic techniques to generate arrays of immobilized oligonucleotide probes (DNA chips). These methods, in which light is used to direct the synthesis of oligonucleotide probes in high-density, miniaturized arrays, utilize photolabile 5'-protected *N*-acyl-deoxynucleoside phosphoramidites, surface linker chemistry and versatile combinatorial synthesis strategies. A matrix of 256 spatially defined oligonucleotide probes may be generated in this manner.

4.21 PREPARATION OF NUCLEIC ACID FRAGMENTS

The nucleic acids may be obtained from any appropriate source, such as cDNAs, genomic DNA, chromosomal DNA, microdissected chromosome bands, cosmid or YAC inserts, and RNA, including mRNA without any amplification steps. For example, Sambrook *et al.* (1989) describes three protocols for the isolation of high molecular weight DNA from mammalian cells (p. 9.14-9.23).

DNA fragments may be prepared as clones in M13, plasmid or lambda vectors and/or prepared directly from genomic DNA or cDNA by PCR or other amplification methods. Samples may be prepared or dispensed in multiwell plates. About 100-1000 ng of DNA samples may be prepared in 2-500 ml of final volume.

The nucleic acids would then be fragmented by any of the methods known to those of skill in the art including, for example, using restriction enzymes as described at 9.24-9.28 of Sambrook *et al.* (1989), shearing by ultrasound and NaOH treatment.

Low pressure shearing is also appropriate, as described by Schriefer *et al.* (1990) Nucleic Acids Res. 18(24) 7455-6, incorporated herein by reference). In this method, DNA samples are passed through a small French pressure cell at a variety of low to intermediate pressures. A lever device allows controlled application of low to intermediate pressures to the cell. The results of these studies indicate that low-pressure shearing is a useful alternative to sonic and enzymatic DNA fragmentation methods.

One particularly suitable way for fragmenting DNA is contemplated to be that using the two base recognition endonuclease, *CviJI*, described by Fitzgerald *et al.* (1992) Nucleic Acids Res. 20(14) 3753-62. These authors described an approach for the rapid fragmentation and fractionation

of DNA into particular sizes that they contemplated to be suitable for shotgun cloning and sequencing.

The restriction endonuclease *CviJI* normally cleaves the recognition sequence PuGCPy between the G and C to leave blunt ends. Atypical reaction conditions, which alter the specificity of this enzyme (*CviJI***), yield a quasi-random distribution of DNA fragments form the small molecule pUC19 (2688 base pairs). Fitzgerald *et al.* (1992) quantitatively evaluated the randomness of this fragmentation strategy, using a *CviJI*** digest of pUC19 that was size fractionated by a rapid gel filtration method and directly ligated, without end repair, to a lac Z minus M13 cloning vector. Sequence analysis of 76 clones showed that *CviJI*** restricts pyGCPy and PuGCPu, in addition to PuGCPy sites, and that new sequence data is accumulated at a rate consistent with random fragmentation.

As reported in the literature, advantages of this approach compared to sonication and agarose gel fractionation include: smaller amounts of DNA are required (0.2-0.5 ug instead of 2-5 ug); and fewer steps are involved (no preligation, end repair, chemical extraction, or agarose gel electrophoresis and elution are needed

Irrespective of the manner in which the nucleic acid fragments are obtained or prepared, it is important to denature the DNA to give single stranded pieces available for hybridization. This is achieved by incubating the DNA solution for 2-5 minutes at 80-90°C. The solution is then cooled quickly to 2°C to prevent renaturation of the DNA fragments before they are contacted with the chip. Phosphate groups must also be removed from genomic DNA by methods known in the art.

4.22 PREPARATION OF DNA ARRAYS

Arrays may be prepared by spotting DNA samples on a support such as a nylon membrane. Spotting may be performed by using arrays of metal pins (the positions of which correspond to an array of wells in a microtiter plate) to repeated by transfer of about 20 nl of a DNA solution to a nylon membrane. By offset printing, a density of dots higher than the density of the wells is achieved. One to 25 dots may be accommodated in 1 mm^2, depending on the type of label used. By avoiding spotting in some preselected number of rows and columns, separate subsets (subarrays) may be formed. Samples in one subarray may be the same genomic segment of DNA (or the same gene) from different individuals, or may be different, overlapped genomic clones. Each of the subarrays may represent replica spotting of the same samples. In one example, a selected gene segment may be amplified from 64 paticnts. For each patient, the amplified gene segment may be in one 96-well plate (all 96 wells containing the same sample). A plate for each of the 64 patients is prepared. By using a 96-pin device, all samples may be spotted on one 8 x 12 cm membrane.

Subarrays may contain 64 samples, one from each patient. Where the 96 subarrays are identical, the dot span may be 1 mm^2 and there may be a 1 mm space between subarrays.

Another approach is to use membranes or plates (available from NUNC, Naperville, Illinois) which may be partitioned by physical spacers e.g. a plastic grid molded over the membrane, the grid being similar to the sort of membrane applied to the bottom of multiwell plates, or hydrophobic strips. A fixed physical spacer is not preferred for imaging by exposure to flat phosphor-storage screens or x-ray films.

The present invention is illustrated in the following examples. Upon consideration of the present disclosure, one of skill in the art will appreciate that many other embodiments and variations may be made in the scope of the present invention. Accordingly, it is intended that the broader aspects of the present invention not be limited to the disclosure of the following examples. The present invention is not to be limited in scope by the exemplified embodiments which are intended as illustrations of single aspects of the invention, and compositions and methods which are functionally equivalent are within the scope of the invention. Indeed, numerous modifications and variations in the practice of the invention are expected to occur to those skilled in the art upon consideration of the present preferred embodiments. Consequently, the only limitations which should be placed upon the scope of the invention are those which appear in the appended claims.

All references cited within the body of the instant specification are hereby incorporated by reference in their entirety.

5. EXAMPLES

5.1 EXAMPLE 1

Novel Nucleic Acid Sequences Obtained From Various Libraries

A plurality of novel nucleic acids were obtained from cDNA libraries prepared from various human tissues and in some cases isolated from a genomic library derived from human chromosome using standard PCR, SBH sequence signature analysis and Sanger sequencing techniques. The inserts of the library were amplified with PCR using primers specific for the vector sequences which flank the inserts. Clones from cDNA libraries were spotted on nylon membrane filters and screened with oligonucleotide probes (e.g., 7-mers) to obtain signature sequences. The clones were clustered into groups of similar or identical sequences. Representative clones were selected for sequencing.

In some cases, the 5' sequence of the amplified inserts was then deduced using a typical Sanger sequencing protocol. PCR products were purified and subjected to fluorescent dye terminator cycle sequencing. Single pass gel sequencing was done using a 377 Applied Biosystems (ABI) sequencer to obtain the novel nucleic acid sequences. In some cases RACE (Random Amplification of cDNA Ends) was performed to further extend the sequence in the 5' direction.

5.2 EXAMPLE 2

Assemblage of Novel Nucleic Acids

The contigs or nucleic acids of the present invention, designated as SEQ ID NO: 345-516 were assembled using an EST sequence as a seed. Then a recursive algorithm was used to extend the seed EST into an extended assemblage, by pulling additional sequences from different databases (i.e., Hyseq's database containing EST sequences, dbEST version 114, gb pri 114, and UniGene version 101) that belong to this assemblage. The algorithm terminated when there was no additional sequences from the above databases that would extend the assemblage. Inclusion of component sequences into the assemblage was based on a BLASTN hit to the extending assemblage with BLAST score greater than 300 and percent identity greater than 95%.

Table 6 sets forth the novel predicted polypeptides (including proteins) encoded by the novel polynucleotides (SEQ ID NO: 517-688) of the present invention, and their corresponding nucleotide locations to each of SEQ ID NO: 517-688. Table 6 also indicates the method by which the polypeptide was predicted. Method A refers to a polypeptide obtained by using a software program called FASTY (available from http://fasta.bioch.virginia.edu) which selects a polypeptide based on a comparison of the translated novel polynucleotide to known polynucleotides (W.R. Pearson, Methods in Enzymology, 183:63-98 (1990), herein incorporated by reference). Method B refers to a polypeptide obtained by using a software program called GenScan for human/vertebrate sequences (available from Stanford University, Office of Technology Licensing) that predicts the polypeptide based on a probabilistic model of gene structure/compositional properties (C. Burge and S. Karlin, J. Mol. Biol., 268:78-94 (1997), incorporated herein by reference). Method C refers to a polypeptide obtained by using a Hyseq proprietary software program that translates the novel polynucleotide and its complementary strand into six possible amino acid sequences (forward and reverse frames) and chooses the polypeptide with the longest open reading frame.

5.3 EXAMPLE 3

Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), full-length gene cDNA sequences and their corresponding protein sequences were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genebank. Other computer programs which may have been used in the editing process were phredPhrap and Consed ((University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.)). The full-length nucleotide sequences are shown in the Sequence Listing as SEQ ID NO: 1-23. The amino acids are SEQ ID NO: 173-195.

Table 1 shows the various tissue sources of SEQ ID NO: 1-23.

The nearest neighbor results for SEQ ID NO: 1-23 were obtained by a BLASTP (version 2.0al 19MP-WashU) search against Genpept release 121 and Geneseq release 200101 (Derwent), using BLAST algorithm. The nearest neighbor result showed the closest homologue for SEQ ID NO: 1-23 from Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologs with identifiable functions for SEQ ID NO: 1-23 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 7 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.4 EXAMPLE 4

Novel Nucleic Acids

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e. dbEST version 117, gb pri 117, UniGene version 117, Genpept release 117). Other computer programs which may have been used

in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 24-137. The corresponding amino acids are SEQ ID NO: 196-309.

Table 1 shows the various tissue sources of SEQ ID NO: 24-137.

The nearest neighbor results for SEQ ID NO: 24-137 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 121 and Geneseq release 200101 (Derwent), using BLAST algorithm. The nearest neighbor result showed the closest homologue for SEQ ID NO: 24-137 from Genpept . The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologs with identifiable functions for SEQ ID NO: 24-137 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, the eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 7 shows the position of the signal peptide in each of the polypeptides and the maximum score and mean score associated with that signal peptide.

5.5 EXAMPLE 5)

<u>Novel Nucleic Acids</u>

Using PHRAP (Univ. of Washington) or CAP4 (Paracel), a full-length gene cDNA sequence and its corresponding protein sequence were generated from the assemblage. Any frame shifts and incorrect stop codons were corrected by hand editing. During editing, the sequence was checked using FASTY and/or BLAST against Genbank (i.e., dbEST version 118, gb pri 118, UniGene version 118, Genpept release 118). Other computer programs which may have been used in the editing process were phredPhrap and Consed (University of Washington) and ed-ready, ed-ext and gc-zip-2 (Hyseq, Inc.). The full-length nucleotide, including splice variants resulting from these procedures are shown in the Sequence Listing as SEQ ID NOS: 138-172. The corresponding amino acid sequences are SEQ ID NO: 310-344.

Table 1 shows the various tissue sources of SEQ ID NO: 138-172.

The homology results for SEQ ID NO: 138-172 were obtained by a BLASTP version 2.0al 19MP-WashU search against Genpept release 121 and Geneseq release 200101 (Derwent), using BLAST algorithm. The nearest neighbor result showed the homologs for SEQ ID NO: 138-172 from Genpept. The translated amino acid sequences for which the nucleic acid sequence encodes are shown in the Sequence Listing. The homologues with identifiable functions for SEQ ID NO: 138-172 are shown in Table 2 below.

Using eMatrix software package (Stanford University, Stanford, CA) (Wu et al., J. Comp. Biol., Vol. 6 pp. 219-235 (1999) herein incorporated by reference), all the sequences were examined to determine whether they had identifiable signature regions. Table 3 shows the signature region found in the indicated polypeptide sequences, the description of the signature, thc eMatrix p-value(s) and the position(s) of the signature within the polypeptide sequence.

Using the pFam software program (Sonnhammer et al., Nucleic Acids Res., Vol. 26(1) pp. 320-322 (1998) herein incorporated by reference) all the polypeptide sequences were examined for domains with homology to certain peptide domains. Table 4 shows the name of the domain found, the description, the p-value and the pFam score for the identified domain within the sequence.

The nucleotide sequence within the sequences that codes for signal peptide sequences and their cleavage sites can be determine from using Neural Network SignalP V1.1 program (from Center for Biological Sequence Analysis, The Technical University of Denmark). The process for identifying prokaryotic and eukaryotic signal peptides and their cleavage sites are also disclosed by Henrik Nielson, Jacob Engelbrecht, Soren Brunak, and Gunnar von Heijne in the publication " Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites" Protein Engineering, Vol. 10, no. 1, pp. 1-6 (1997), incorporated herein by reference. A maximum S score and a mean S score, as described in the Nielson et as reference, was obtained for the polypeptide sequences. Table 7 shows the position of the signal peptide in

each of the polypeptides and the maximum score and mean score associated with that signal peptide.

Table 5 is a correlation table of all of the sequences and SEQ ID Nos.

TABLE 1

Tissue Origin	RNA Source	Library Name	SEQ ID NOS:
fibroblast	Strategene	LFB001	37 62-63 69 87-89 108 124 126-127 144-145 166 168
NULL	NULL	CGSd001	25 54 74
NULL	NULL	CGSd004	26 80
NULL	NULL	CGSd005	11 91 166
NULL	NULL	CGSd006	11 25-26 34 39 60 139 143 160 164 166
NULL	NULL	CGSd009	13 45
NULL	NULL	CGd007	34
NULL	NULL	CGd008	34 39 76
NULL	NULL	CTL016	170
NULL	NULL	CTL021	57 76 78
NULL	NULL	PCR2V1	1 57 95
NULL	NULL	PGEMV1	10 26 32 34 42 44 52 57 62 70 78 95 113-114 122-124 144-145 150
NULL	NULL	SUP002	32 35 57 62 78 149 153-154 168
NULL	NULL	TEST	52
adipocytes	Stratagene	ADP001	6 8 38 91 95 97 104 108 110 144-145 153-154 167
adrenal gland	Clontech	ADR002	6 9 14 16 22 27 38-39 41 45 52 54 62-63 74 79 81 90 107 111 123 126 128 136 166 168
adult brain	BioChain	ABR012	38 126
adult brain	BioChain	ABR013	35 53
adult brain	Clontech	ABR001	15 17 22 35 72 84 126
adult brain	Clontech	ABR006	11 22-23 31-32 39 43 52 54-55 60 81 84 102 110 139 149 168
adult brain	Clontech	ABR008	1-4 7-12 18 22-24 27-28 31-33 35 40 44-45 50 52-53 55 57-60 62 72 74 81 84 87-89 93-96 98-99 102 104 107-108 110-111 113-114 120-123 130-131 136 142-145 147-148 150 157 160 164 167-168 171-172
adult brain	Clontech	ABR011	23 38
adult brain	GIBCO	AB3001	6 17 19 22 38 52 54-55 62 74 86 104 122 127-128 131 164
adult brain	GIBCO	ABD003	6 10 18 22-23 32 35 38 40 52-53 55 58 62 71-72 74 79-81 94-95 99 105 108 111 113-114 122 126 128 130-131 139 142 160-161 167-169
adult brain	Invitrogen	ABR014	57 62 76 78 96 120-121 161
adult brain	Invitrogen	ABR015	6 37 76 91 99
adult brain	Invitrogen	ABR016	57 63
adult brain	Invitrogen	ABT004	10 23 32 40 43 55 58 60 74 84 99 112 120-121 148 167-168

Tissue Origin	RNA Source	Library Name	SEQ ID NOS:
adult heart	GIBCO	AHR001	2 6 14 18-19 22 30 38-39 45 51-54 57 60 62-63 66-67 74 76 79 81 87-91 93-95 98 104 108 110 113-114 122-124 126-128 135-136 140-141 144-145 150 153-154 160 163 168-169 171-172
adult kidney	GIBCO	AKD001	2 6-10 14 17-19 22 25 28-29 33-35 38-39 42-45 52 54-55 58 60-63 67 72 74 79-80 86-89 91-96 99-101 104-105 107-108 111-114 116 122-124 126-128 130-132 136 138-139 141 144-145 147 155 160 167-168 172
adult kidney	Invitrogen	AKT002	1-2 6 20 34 38 41 43 52 62 76 81 87-90 94-95 105 138 152-154 160
adult liver	Clontech	ALV003	38 57 76 95 149
adult liver	Invitrogen	ALV002	9 14 22 26 41-42 45 55 62 65 87-89 97 111 120-121 141 149 160 162 164 168
adult lung	GIBCO	ALG001	38 41 54-55 57 62 76 79 81 86 91 104 111-112 120-121 123 160 163
adult lung	Invitrogen	LGT002	1-3 5 8 10 14 17 19 21-23 25 27 29 34 37-38 40-41 44-45 52-54 60-62 68-70 81 84 86-91 93-96 99-101 104 106-108 112-114 122-123 127 131-132 135-136 141 144-145 148 152-157 160-162 165 167-168 171-172
adult spleen	Clontech	SPLc01	9 16 52 55 62 81 143 150 153-154
adult spleen	GIBCO	ASP001	1-2 6 10 18 34 38 40 42 51 57 61 72 74-76 78 80-81 86 90-91 95 104 108 110-111 122-123 126 130-131 143 153-154 168 172
bladder	Invitrogen	BLD001	1 15 18 34 51 87-89 95 97 132 164
bone marrow	Clonetech	BMD007	57 76 138
bone marrow	Clontech	BMD001	1-3 6 8-11 13-14 19 21 31 36 38-41 50 53-54 57 60-62 69 74 76 78 81 85 90-91 93-96 105 107-109 112-114 122 125-128 139 141 143 148 153-154 159 161-162 166 169-172
bone marrow	Clontech	BMD004	38 57 76 91 98
bone marrow	GF	BMD002	9 12 16 19 23 25 27 39 41-44 52-59 62 73-74 76 78 81 84 91 94-99 109-110 113-114 120-121 125-128 131 135 137 139 141-143 150 153-154 162 168 172
cervix	BioChain	CVX001	2 6-7 10 17 19-20 23-24 27 34 38 52 54 60 62-63 69 80 87-91 93-96 100-101 108 111 113-114 123 127-128 134 136 141 144-145 152-154

Tissue Origin	RNA Source	Library Name	SEQ ID NOS:
			158 162 165
colon	Invitrogen	CLN001	23-24 55 65 72 95 120-123 142 171
diaphragm	BioChain	DIA002	62 126 153-154
endothelial cells	Strategene	EDT001	1 6-7 10-11 14 16-19 22 24 27 35 37-40 42-43 47 49 52 58 60 62-63 67 69 72 74 81 90-91 93-101 104 107-108 120-124 126-133 135-136 141-142 144-145 151 153-154 156 160 165 167-168 170-172
esophagus	BioChain	ESO002	23
fetal brain	Clontech	FBR001	60 91
fetal brain	Clontech	FBR004	1 3 11 22 32 74
fetal brain	Clontech	FBR006	1 9-12 14 16 22 29 32-33 36 40 42 45 57 59-60 72 80-81 84 91 94 97-98 108 111 113-114 119-123 135-137 143-145 147-148 157 160 166 169 171-172
fetal brain	GIBCO	HFB001	1-2 6-7 10 13 22-23 27-28 32 37-39 42 44 48 51-52 54 57-58 60 62 69 72 76 78-80 91 93-95 102 108 112-114 120-123 126 128 130-132 135 138-139 144-145 152 155 160-161 164 167
fetal brain	Invitrogen	FBT002	1 3 8 22-23 40 42 62 74 78 80 84 91 95 98 100-101 108 122 124 136 142-145 147 160
fetal heart	Invitrogen	FHR001	57 160
fetal kidney	Clontech	FKD001	7 38 40 60 62 85 87-89 95 147
fetal kidney	Clontech	FKD002	42 95
fetal kidney	Invitrogen	FKD007	57 62 78 91 95
fetal liver	Clontech	FLV004	41 44 57 62 69 72 95 111 149 169
fetal liver	Invitrogen	FLV001	3 18 22 27 34 52 55 57-58 76 78 87-89 91 111 120-121 123 131 136 144-145 149 168 170
fetal liver-spleen	Soares	FLS001	1-3 5-6 10-11 13-14 16-19 23-24 26-27 29 31 34-35 38-44 46 50-52 54 57-65 67-70 72 74 76 78-81 84 86-91 93-98 103-104 106 108 111-112 117 120-124 126-128 130-133 135-136 138 141 143-145 147 149 155-156 160-162 165 167-172
fetal liver-spleen	Soares	FLS002	3 5-6 8-10 13 16-19 23-24 26-27 34-35 37 41 44-45 50-51 53-55 57-58 62-64 68-70 74-75 79 81 84 87-89 91 93-94 96 98-99 103-105 108 111-115 117 119 123-124 126-128 131-132 134-136 143-148 150 153-154 156 161-162 164 168 170-172

Tissue Origin	RNA Source	Library Name	SEQ ID NOS:
fetal liver-spleen	Soares	FLS003	26 50 53 57 62 76 78 96 149-150
fetal lung	Clontech	FLG001	94-95 115 131 149 168
fetal lung	Invitrogen	FLG003	1 24 40 57 62 73 76 78 80-81 95 111 136 153-154 168
fetal muscle	Invitrogen	FMS001	24 27 30 54-58 90 94 99 119 123 131 140 142 147 164 168
fetal muscle	Invitrogen	FMS002	42 122 155
fetal skin	Invitrogen	FSK001	1 6 18 22-24 27 43 45 52 55 57 60 62 68-69 78-79 81 91 94-95 97 99 108 112-114 120-121 142 147 153-154 161 164-165 167-170
fetal skin	Invitrogen	FSK002	1 29 40-41 54-55 62 69 94 96 115 153-154
fetal spleen	BioChain	FSP001	91 95
infant brain	NULL	IBM002	3 11 23 32 128
infant brain	NULL	IBS001	44 59 120-121 170
infant brain	Soares	IB2002	3 10-13 18-19 22-23 25 27 32 37-40 43 54-55 57 59-63 77-78 84 94 98-99 102 105 108 120-121 123-124 127 130 136-137 143-145 152 155-156 160 164 167-168 170
infant brain	Soares	IB2003	1 3 23 27 37 40 51 60 62 72 99 104 108 111 120-123 128 136 139 143-145 167
leukocytes	Clontech	LUC003	1-2 38 62 76 84 94 97 128 136 161 165 172
leukocytes	GIBCO	LUC001	2 6-8 11 13-14 16 18-19 22-24 27-28 37-38 44-45 52-53 57-58 60 62 67 69-70 74 76 78-81 84 90-91 93-95 97-101 107-114 122-123 126-128 130-132 135-136 138-139 141 143 147 156-157 161-162 165-167
lung	37 62-63 69 87-89 108 124 126-127 144-145 166 168		
lung	Strategene	LFB001	1-2 6-7 10 14 19
lymph node	Clontech	ALN001	6 19 34 39 62 80 90 93 95 112-114 127 143 162
lymphocyte	ATCC	LPC001	37-38 43 51 55 57 62 70 72 76 78-79 90 94-95 97 108 112 119 122 132 143 146 165-166
macrophage	Invitrogen	HMP001	113-114
mammary gland	Invitrogen	MMG001	1-2 8 13-14 18 22-23 27 33 40 43 52 55 58-60 70 74 84 86-90 93 95 97 99-101 104 107-108 111 113-114 120-121 123 131 136 142-145 147

Tissue Origin	RNA Source	Library Name	SEQ ID NOS:
			110 123 144-146 158
thymus	Clonetech	THM001	6 9 18-19 34 38 41-42 45 52-54 63 69 84 94-96 99 112 127-128 130-131 141 160 170 172
thymus	Clontech	THMc02	1 3 7 9 16 23 27 35 40 44 51-53 62-63 72 81 84 86-90 94-95 99-101 115 123 128 132 141 143-145 148 167 169-171
thyroid gland	Clontech	THR001	1 3 6 16-19 22 29 35 38 42 45 52 54 56 58 60-63 74 76 81 84 87-89 91 93-95 97 99-101 105 107-108 111 113-114 122-123 126-128 131 135-136 141 144-145 147 160-162 164-165 168 172
trachea	Clontech	TRC001	5 69 76 80 90-91 93 113-114 166
umbilical cord	BioChain	FUC001	6-7 10-11 14 22 29 41 43 51 57 62 74 76 78 80-81 91 94 97 105 110 115 124 131-132 141 144-145 150 160 165 168-169 171-172
uterus	Clontech	UTR001	37 52 67 76 81 95 108 112 118 142 168
young liver	GIBCO	ALV001	2 26 29 38 52 54 63 72 74 95-97 99 104 108 111 120-121 123 149 156 164 169 172

TABLE 2

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
1	W80293	*Homo sapiens*	*Human translocation* associated protein designated Gp25L-H.	1016	96
2	M74089	Homo sapiens	TB1	2226	99
3	D88827	Homo sapiens	zinc finger protein FPM315	447	46
4	Z34818	Homo sapiens	voltage-dependent L-type Ca channel alpha 1 subunit	1363	96
5	U60269	Homo sapiens	putative envelope protein; orf similar to env of Type A and Type B retroviruses and to class II HERVs	370	98
6	Z47087	Homo sapiens	RNA polymerase II elongation factor-like protein	865	100
7	AL078461	Homo sapiens	dJ901O8.1 (continues in dJ1121G12 (AL109965))	782	97
8	X93357	Mus musculus	homolog of human SYT	870	53
9	AF227156	Homo sapiens	RRN3	1679	98
10	AC004908	Homo sapiens	similar to ribosomal protein L23a; similar to P29316 (PID:g132848)	377	94
11	U10556	Saccharomyces cerevisiae	Yhr085wp	175	31
12	M29581	Homo sapiens	zinc finger protein 8 (ZFP8)	2949	99
13	M64788	Homo sapiens	GTPase activating protein	1235	55
14	AK026016	Homo sapiens	unnamed protein product	1846	100
15	G00637	Homo sapiens	Human secreted protein, SEQ ID NO: 4718.	166	64
16	AF201941	Homo sapiens	DC8	1226	100
17	Z83826	Homo sapiens	dJ473B4.1 (novel protein similar to predicted human and worm genes)	1083	100
18	AB031052	Homo sapiens	vaccinia related kinase 3	2521	100
19	AL050091	Homo sapiens	hypothetical protein	1011	100
20	AF090931	Homo sapiens	PRO0483	137	65
21	U39320	Rattus norvegicus	cysteine string protein	747	68
22	Y66718	Homo sapiens	Membrane-bound protein PRO1106.	1554	70
24	AE003612	Drosophila melanogaster	CG12393 gene product	244	36
25	V00662	Homo sapiens	ATPase 6	232	100
26	T12920_cd1	Homo sapiens	11-AUG-1994 Human serum albumin gene with restriction site between domains 2 and 3.	731	100
27	AL109804	Homo sapiens	dJ1009E24.3 (A novel protein)	920	100
28	AL021918	Homo sapiens	b34I8.1 (Kruppel related Zinc Finger protein 184)	649	54
29	G01131	Homo sapiens	Human secreted protein, SEQ ID NO: 5212.	666	99
30	Y73373	Homo sapiens	HTRM clone 921803 protein sequence.	1225	100
31	AJ293573	Homo sapiens	zinc finger protein Cezanne	2243	100

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
32	AF157634	Homo sapiens	collapsin response mediator protein-5	2943	99
33	J00287	Homo sapiens	pepsinogen	2032	100
34	AF264014	Homo sapiens	scavenger receptor cysteine-rich type 1 protein M160 precursor	4523	99
35	AK024433	Homo sapiens	FLJ00023 protein	918	100
36	AF119855	Homo sapiens	PRO1847	253	71
37	AF176667	Xenopus laevis	F-box protein 28	426	80
38	M85148	Macaca mulatta	cytochrome oxidase subunit III	178	68
39	R83119	Homo sapiens	Human cisplatin resistance protein.	396	98
40	X58521	Homo sapiens	nucleoporin p62	2068	99
41	Y86211	Homo sapiens	Nuclear transport protein clone hfb066 protein sequence.	1432	87
42	AL035587	Homo sapiens	dJ475N16.3 (novel protein similar to RPL7A (60S ribosomal protein L7A))	1243	99
43	AF233522	Homo sapiens	gamma-adaptin related protein, GGA2	3196	100
44	AF151066	Homo sapiens	HSPC232	1513	93
45	AF113539	Homo sapiens	hypothalamus protein HT001	1647	99
46	AF090901	Homo sapiens	PRO0195	120	60
47	J02459	bacteriophage lambda	D (head-DNA stabilization;110)	561	99
48	AF068294	Homo sapiens	HDCMB45P	234	72
49	J02459	bacteriophage lambda	R (cell lysis;158)	834	100
50	U71363	Homo sapiens	zinc finger protein zfp6	396	47
51	AL022393	Homo sapiens	p373c6.1	2751	99
52	AF257330	Homo sapiens	COBW-like protein	1188	99
53	Y07010	Homo sapiens	Breast cancer associated antigen precursor sequence.	767	98
54	AL158056	Schizosaccharomyces pombe	similar to yeast Ecm29 cell wall stucture/biosynthesis protein	813	25
55	Y44559	Homo sapiens	Human Rhotekin protein.	2721	100
56	D86639	Mus musculus	stac	360	38
57	V00488	Homo sapiens	alpha globin	382	98
58	AL033378	Homo sapiens	dJ323M4.1 (KIAA0790 protein)	6030	100
59	AB026111	Homo sapiens	XAB2	4470	99
60	AB043550	Mus musculus	neural activity-related ring finger protein	3852	99
61	D38112	Homo sapiens	ATPase subunit 6	260	98
62	X15187	Homo sapiens	precursor polypeptide (AA -21 to 782)	4112	100
63	D50369	Homo sapiens	low molecular mass ubiquinone-binding protein	383	98
64	AF161356	Homo sapiens	HSPC093	126	58
65	Z11502	Homo sapiens	intestine-specific annexin	1426	99
66	J03941	Mus musculus	ferritin heavy chain	962	100
67	AB034206	Homo sapiens	MCT-1	962	100
68	AF118086	Homo sapiens	PRO1992	107	58
69	AF182416	Homo sapiens	MDS015	1739	97
70	G03240	Homo sapiens	Human secreted protein,	108	56

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
			SEQ ID NO: 7321.		
71	G00397	Homo sapiens	Human secreted protein, SEQ ID NO: 4478.	161	62
72	X76538	Homo sapiens	hMpv17	569	100
73	AC005626	Homo sapiens	R29124_1	1199	97
74	G01332	Homo sapiens	Human secreted protein, SEQ ID NO: 5413.	407	87
75	Y53883	Homo sapiens	A suppressor of cytokine signalling protein designated HSCOP-3.	1250	100
76	M34059	Homo sapiens	beta-globin	962	100
78	V00488	Homo sapiens	alpha globin	733	100
79	Z11907_cd1	Homo sapiens	19-JAN-1999 Human potassium channel K+Hnov28 cDNA (5' splice variant 1).	1282	100
80	U47924	Homo sapiens	C10	627	100
81	AF134894	Homo sapiens	SWAP-70 homolog	3019	100
82	AB042027	Mus musculus	Golgi-associated band 4.1-like protein	4806	93
83	G03793	Homo sapiens	Human secreted protein, SEQ ID NO: 7874.	209	71
84	X83301	Homo sapiens	SMA5	754	97
85	V00662	Homo sapiens	URF 2 (NADH dehydrogenase subunit)	211	91
86	W88627	Homo sapiens	Secreted protein encoded by gene 94 clone HPMBQ32.	230	69
87	Y27607	Homo sapiens	Human secreted protein encoded by gene No. 41.	1461	100
88	Y27607	Homo sapiens	Human secreted protein encoded by gene No. 41.	1340	94
89	Y94295	Homo sapiens	Human coenzyme A-utilising enzyme CoAEN-3.	1068	100
90	AL080169	Homo sapiens	hypothetical protein	973	98
91	D14531	Homo sapiens	'human homologue of rat ribosomal protein L9'	974	100
93	AC006465	Homo sapiens	supported by mouse EST AA277724 (NID:g1917632) and Genscan	437	100
94	AL137798	Homo sapiens	dJ1182A14.5.1 (novel gene (isoform 1))	2075	94
95	M77232	Homo sapiens	ribosomal protein S6	706	97
96	AF261717	Homo sapiens	SAR1	1032	100
97	X06256	Homo sapiens	integrin alpha 5 subunit precursor	5541	100
98	AC004472	Homo sapiens	P1.11659_5	542	98
99	W85459	Homo sapiens	Secreted protein encoded by clone dh1135_9.	1650	98
100	AF201333	Homo sapiens	delta-tubulin	2380	100
101	AF201333	Homo sapiens	delta-tubulin	1345	100
102	AB036836	Homo sapiens	Carbonic anhydrase-related protein 10	1740	100
104	W80748	Homo sapiens	Human mitochondrial chaperone protein (Hmt-GrpE).	1102	100
105	G00376	Homo sapiens	Human secreted protein, SEQ ID NO: 4457.	130	73
106	Y99444	Homo sapiens	Human PRO1575 (UNQ781) amino acid	1241	94

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
			sequence SEQ ID NO:358.		
107	G03787	Homo sapiens	Human secreted protein, SEQ ID NO: 7868.	232	74
108	AJ237734	Homo sapiens	ribophorin II	2619	100
110	Y99349	Homo sapiens	Human PRO1110 (UNQ553) amino acid sequence SEQ ID NO:31.	1683	100
111	AF258341	Homo sapiens	NADPH-cytochrome P450 reductase	3540	99
112	AB044547	Homo sapiens	heterogeneous nuclear ribonucleoprotein L	3178	99
113	AF067804	Homo sapiens	HDCMC04P	2124	94
114	AF067804	Homo sapiens	HDCMC04P	2219	97
115	U73778	Homo sapiens	collagen type XII alpha-1	16043	99
116	G02922	Homo sapiens	Human secreted protein, SEQ ID NO: 7003.	120	69
118	G03133	Homo sapiens	Human secreted protein, SEQ ID NO: 7214.	175	59
119	AF115345	Homo sapiens	calcium-regulated heat stable protein CRHSP-24	282	100
120	AF265228	Homo sapiens	DNMT1 associated protein-1	2422	100
121	AF265228	Homo sapiens	DNMT1 associated protein-1	1320	85
122	AB018189	Xenopus laevis	NO27	907	76
123	AB033603	Homo sapiens	cop9 complex subunit 7a	1364	100
124	AL121805	Drosophila melanogaster	BACN4L24.f	200	32
125	Y03000	Homo sapiens	Fragment of human secreted protein encoded by gene 121.	153	72
126	D38112	Homo sapiens	NADH dehydrogenase subunit 4L	448	90
127	AL050091	Homo sapiens	hypothetical protein	1011	100
128	AK023702	Homo sapiens	unnamed protein product	2774	100
129	J02459	bacteriophage lambda	M (tail component;109)	578	100
130	AF116272	Homo sapiens	T-cell activation protein	532	85
131	W74836	Homo sapiens	Human secreted protein encoded by gene 108 clone HEBEK93.	655	100
132	AF146793	Mus musculus	PDCL2	705	58
133	J02459	bacteriophage lambda	W (head-tail joining;68)	336	100
134	AB011539	Homo sapiens	MEGF6	131	58
135	G00827	Homo sapiens	Human secreted protein, SEQ ID NO: 4908.	523	98
136	Y35969	Homo sapiens	Extended human secreted protein sequence, SEQ ID NO. 218.	1777	99
137	AB007861	Homo sapiens	KIAA0401	1856	100
138	AK000529	Homo sapiens	unnamed protein product	1027	91
139	A17783	unidentified	NC28	583	100
140	Y73373	Homo sapiens	HTRM clone 921803 protein sequence.	561	100
141	D00762	Homo sapiens	proteasome subunit C8	686	93
142	X76091	Homo sapiens	DNA binding protein RFX2	3747	99
143	AK000004	Homo sapiens	FLJ00004 protein	3106	96
144	Y73377	Homo sapiens	HTRM clone 1645941	852	41

SEQ ID NO:	ACCESSION NUMBER	SPECIES	DESCRIPTION	SMITH-WATERMAN SCORE	% IDENTITY
			protein sequence.		
145	Y73377	Homo sapiens	HTRM clone 1645941 protein sequence.	591	45
146	U12329	Cricetulus griseus	mutant sterol regulatory element binding protein-2	373	94
147	U80735	Homo sapiens	CAGF28	3779	97
148	Y91577	Homo sapiens	Human secreted protein sequence encoded by gene 2 SEQ ID NO:250.	155	96
149	X00129	Homo sapiens	precursor RBP	1005	96
150	X97675	Homo sapiens	plakophilin 2a	4288	99
151	J02459	bacteriophage lambda	I (tail component;223)	1111	99
152	Y94702	Homo sapiens	Human androgen shutoff gene 3 (AS3) protein sequence.	839	100
153	X95325	Homo sapiens	DNA-binding protein	1975	100
154	AF171061	Canis familiaris	Y-box protein ZONAB-A	1426	89
155	X99584	Homo sapiens	SMT3A protein	522	98
156	Y99397	Homo sapiens	Human PRO1298 (UNQ666) amino acid sequence SEQ ID NO:210.	1554	100
157	X78926	Homo sapiens	zinc finger protein	426	50
158	J02459	bacteriophage lambda	E (capsid component;341)	1754	100
159	G02922	Homo sapiens	Human secreted protein, SEQ ID NO: 7003.	160	48
160	AF157317	Homo sapiens	AD-015 protein	869	100
161	Y36495	Homo sapiens	Fragment of human secreted protein encoded by gene 27.	111	67
162	Y94870	Homo sapiens	Human protein clone HP02545.	1707	100
163	AF204172	Homo sapiens	popeye protein 1	1872	98
164	B01391	Homo sapiens	Neuron-associated protein.	376	58
165	Z99259	Schizosaccharomyces pombe	putative phosphotransferase	396	43
166	U28831	Homo sapiens	protein that is immuno-reactive with anti-PTH polyclonal antibodies	518	55
167	AF225419	Homo sapiens	HSCARG	1546	100
168	X04412	Homo sapiens	plasma gelsolin	4101	100
169	X13482	Homo sapiens	U2 snRNP-specific A' protein (AA 1-255)	1284	99
170	X59618	Homo sapiens	small subunit ribonucleotide reductase	1389	100
171	AC004882	Homo sapiens	similar to cytochrome Bc1 J chain; similar to 1BGY (PID:g4139401)	326	100
172	AL034374	Homo sapiens	dJ483K16.1.1 (novel protein (isoform 1))	1651	100

TABLE 3

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	RESULTS
1	PF01105	emp24/gp25L/p24 family.	PF01105B 25.12 1.257e-39 170-222
2	BL00215	Mitochondrial energy transfer proteins.	BL00215B 10.44 1.900e-09 481-494
3	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 6.400e-16 427-440 PD00066 13.92 7.923e-15 371-384 PD00066 13.92 4.600e-14 343-356 PD00066 13.92 3.500e-13 399-412
4	PR00167	CALCIUM CHANNEL SIGNATURE	PR00167D 13.86 3.400e-34 374-401 PR00167E 20.54 2.800e-27 1018-1039 PR00167F 16.75 4.000e-20 1551-1566 PR00167G 10.89 5.050e-16 1599-1612 PR00167D 13.86 3.778e-12 1599-1626
6	BL00517	Ribonuclease III family proteins.	BL00517C 7.29 9.386e-10 98-112
8	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 5.500e-09 325-344
10	PR00925	NONHISTONE CHROMOSOMAL PROTEIN HMG17 FAMILY SIGNATURE	PR00925B 3.73 3.089e-10 11-24
12	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 8.953e-31 27-66
14	PF00521	DNA gyrase/topoisomerase IV, subunit A.	PF00521G 18.47 7.904e-09 58-92
18	BL00107	Protein kinases ATP-binding region proteins.	BL00107A 18.39 7.207e-11 296-327
21	PR00625	DNAJ PROTEIN FAMILY SIGNATURE	PR00625A 12.84 2.750e-19 30-50 PR00625B 13.48 4.490e-15 61-82
22	PR00927	ADENINE NUCLEOTIDE TRANSLOCATOR 1 SIGNATURE	PR00927B 14.66 6.262e-11 315-337 PR00927E 14.93 3.160e-09 313-335
26	PR00802	SERUM ALBUMIN FAMILY SIGNATURE	PR00802C 12.28 2.370e-11 106-124 PR00802G 14.57 8.734e-09 92-115
28	PR00048	C2H2-TYPE ZINC FINGER SIGNATURE	PR00048A 10.52 5.500e-13 64-78 PR00048A 10.52 7.429e-13 149-163 PR00048A 10.52 1.000e-11 35-49 PR00048B 6.02 6.538e-11 137-147 PR00048A 10.52 6.684e-11 121-135 PR00048B 6.02 2.421e-09 80-90
30	PR00988	URIDINE KINASE SIGNATURE	PR00988A 6.39 8.905e-09 2-20
31	BL00315	Dehydrins proteins.	BL00315A 9.35 8.119e-09 41-69
33	PR00792	PEPSIN (A1) ASPARTIC PROTEASE FAMILY SIGNATURE	PR00792A 11.54 6.250e-23 82-103 PR00792D 12.74 5.200e-18 361-377 PR00792B 12.78 7.750e-14 225-239 PR00792C 9.10 1.000e-12 274-286
34	BL00420	Speract receptor repeat proteins domain proteins.	BL00420B 22.67 3.089e-38 408-463 BL00420B 22.67 8.875e-38 618-673 BL00420B 22.67 6.143e-24 272-327 BL00420B 22.67 7.692e-23 167-222

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	RESULTS
			BL00420B 22.67 3.778e-21 62-117 BL00420B 22.67 7.253e-15 515-570 BL00420C 11.90 1.900e-12 493-504 BL00420C 11.90 3.700e-12 703-714 BL00420C 11.90 9.250e-12 40-51
39	BL00667	Respiratory-chain NADH dehydrogenase subunit 1 proteins.	BL00667A 20.80 1.000e-40 22-74
41	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 8.703e-10 407-461 BL01160B 19.54 2.373e-09 414-468
42	BL00634	Ribosomal protein L30 proteins.	BL00634 34.38 7.796e-20 100-151
50	PD01066	PROTEIN ZINC FINGER ZINC-FINGER METAL-BINDING NU.	PD01066 19.43 9.842e-24 16-55
51	PD00066	PROTEIN ZINC-FINGER METAL-BINDI.	PD00066 13.92 8.200e-16 362-375 PD00066 13.92 4.462e-15 334-347 PD00066 13.92 8.615e-15 473-486 PD00066 13.92 5.200e-14 306-319 PD00066 13.92 3.000e-13 390-403
53	BL01160	Kinesin light chain repeat proteins.	BL01160B 19.54 8.093e-09 586-640
54	PF00956	Nucleosome assembly protein (NAP).	PF00956B 23.14 2.534e-09 896-937
56	BL00479	Phorbol esters / diacylglycerol binding domain proteins.	BL00479A 19.86 6.400e-13 90-113 BL00479B 12.57 5.860e-10 116-132
57	BL01033	Globins profile.	BL01033A 16.94 3.250e-20 26-48
58	PR00910	LUTEOVIRUS ORF6 PROTEIN SIGNATURE	PR00910A 2.51 4.321e-09 9-22
60	BL00518	Zinc finger, C3HC4 type (RING finger), proteins.	BL00518 12.23 9.143e-10 65-74
61	BL00449	ATP synthase a subunit proteins.	BL00449C 13.17 3.813e-21 32-55
62	BL00298	Heat shock hsp90 proteins family proteins.	BL00298A 10.97 1.000e-40 74-119 BL00298E 27.30 1.000e-40 321-376 BL00298F 11.21 1.000e-40 409-464 BL00298H 20.50 1.000e-40 553-607 BL00298C 16.40 2.286e-40 186-230 BL00298B 15.64 1.290e-39 134-181 BL00298G 24.57 5.345e-39 465-520 BL00298I 30.07 7.818e-34 661-715 BL00298D 17.97 6.226e-33 242-282
65	BL00223	Annexins repeat proteins domain proteins.	BL00223C 24.79 1.000e-40 208-263 BL00223B 28.47 9.679e-39 131-181 BL00223A 15.59 6.000e-27 62-96 BL00223C 24.79 1.000e-16 49-104 BL00223A 15.59 2.306e-16 221-255
66	BL00540	Ferritin iron-binding regions proteins.	BL00540A 15.06 1.000e-40 9-50 BL00540B 18.82 1.000e-40 100-155 BL00540C 13.00 7.500e-15 165-177
73	DM00372	CARCINOEMBRYONIC ANTIGEN PRECURSOR AMINO-TERMINAL DOMAIN.	DM00372A 19.18 2.800e-22 19-64 DM00372B 20.31 2.523e-19 83-128

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	RESULTS
74	PD01168	SYNTHETASE LIGASE PROTEIN ALANYL.	PD01168L 9.47 4.500e-09 369-384
76	BL01033	Globins profile.	BL01033A 16.94 7.923e-18 25-47 BL01033B 13.81 1.000e-15 93-105
78	BL01033	Globins profile.	BL01033A 16.94 3.250e-20 283-305 BL01033B 13.81 7.000e-14 345-357
79	PR00169	POTASSIUM CHANNEL SIGNATURE	PR00169A 16.77 2.075e-11 55-75
81	PF00992	Troponin.	PF00992A 16.67 6.921e-09 417-452
82	PR00935	BAND 4.1 PROTEIN FAMILY SIGNATURE	PR00935D 10.20 7.833e-09 240-257
87	BL00166	Enoyl-CoA hydratase/isomerase proteins.	BL00166D 22.87 7.250e-20 177-213 BL00166C 18.93 2.537e-19 126-153 BL00166B 16.92 5.500e-09 80-102
88	BL00166	Enoyl-CoA hydratase/isomerase proteins.	BL00166D 22.87 7.250e-20 177-213 BL00166C 18.93 2.537e-19 126-153 BL00166B 16.92 5.500e-09 80-102
89	BL00166	Enoyl-CoA hydratase/isomerase proteins.	BL00166D 22.87 7.250e-20 146-182 BL00166B 16.92 5.500e-09 80-102
91	BL00700	Ribosomal protein L6 proteins 2.	BL00700B 18.26 1.000e-40 162-201 BL00700D 21.51 1.000e-40 241-285 BL00700A 19.34 9.308e-33 108-146 BL00700C 11.76 4.000e-13 212-222
95	BL00578	Ribosomal protein S6e proteins.	BL00578B 15.15 1.000e-40 39-90 BL00578A 14.24 2.800e-37 1-32
96	BL01020	SAR1 family proteins.	BL01020C 15.35 1.000e-40 83-134 BL01020D 19.09 8.071e-40 159-198 BL01020B 11.70 6.750e-38 41-76 BL01020A 11.87 1.400e-28 7-38
97	BL00242	Integrins alpha chain proteins.	BL00242C 16.86 4.176e-30 319-349 BL00242E 9.03 3.864e-27 1000-1029 BL00242D 13.57 4.706e-26 391-416 BL00242D 13.57 7.407e-13 455-480 BL00242A 13.80 2.500e-11 92-104 BL00242B 8.13 4.913e-11 222-232 BL00242D 13.57 3.700e-10 324-349
100	BL00227	Tubulin subunits alpha, beta, and gamma proteins.	BL00227C 25.48 2.525e-23 114-166 BL00227F 21.16 5.610e-12 400-454
101	BL00227	Tubulin subunits alpha, beta, and gamma proteins.	BL00227C 25.48 2.525e-23 114-166 BL00227F 21.16 5.610e-12 298-352
102	BL00162	Eukaryotic-type carbonic anhydrases proteins.	BL00162C 17.78 1.000e-19 126-163 BL00162E 14.93 5.922e-16 231-264 BL00162D 15.06 7.864e-12 165-190 BL00162F 22.68 7.750e-11 268-302 BL00162A 22.92 3.282e-10 54-85
104	BL01071	grpE protein.	BL01071A 24.88 2.674e-23 76-122 BL01071B 18.21 6.447e-16 191-215
106	PR00002	BONE MATRIX GLA DOMAIN SIGNATURE	PR00002A 11.56 1.750e-16 481-498 PR00002B 8.36 7.632e-13 500-511
111	PR00369	FLAVODOXIN SIGNATURE	PR00369D 20.20 7.261e-16 192-212 PR00369A 11.29 7.600e-14 84-98 PR00369C 11.27 8.313e-12 168-179 PR00369B 13.29 1.600e-11 137-149
112	PD02784	PROTEIN NUCLEAR	PD02784B 26.46 4.713e-16 227-270

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	RESULTS
		RIBONUCLEOPROTEIN.	
113	PF00856	SET domain proteins.	PF00856B 16.42 3.302e-17 382-404 PF00856A 26.14 3.714e-15 129-166
114	PF00856	SET domain proteins.	PF00856B 16.42 3.302e-17 395-417 PF00856A 26.14 3.714e-15 129-166
115	PR00453	VON WILLEBRAND FACTOR TYPE A DOMAIN SIGNATURE	PR00453A 12.79 5.737e-17 1198-1216 PR00453A 12.79 3.571e-16 439-457 PR00453A 12.79 5.125e-15 139-157 PR00453A 12.79 1.563e-13 2322-2340 PR00453B 14.65 7.667e-13 178-193 PR00453B 14.65 7.818e-12 1237-1252 PR00453B 14.65 9.455e-12 2362-2377 PR00453B 14.65 1.243e-11 478-493 PR00453C 12.26 4.462e-10 2428-2437 PR00453C 12.26 9.308e-10 544-553
122	DM00668	ZEIN.	DM00668C 8.89 8.794e-09 24-48
126	PF00420	NADH-ubiquinone/plastoquinone oxidoreductase chain 4L.	PF00420B 16.51 4.150e-31 61-97 PF00420A 16.63 3.571e-24 12-43
128	BL00405	43 Kd postsynaptic protein.	BL00405G 7.78 9.919e-10 285-322
134	PR00011	TYPE III EGF-LIKE SIGNATURE	PR00011D 14.03 2.108e-10 18-37 PR00011B 13.08 6.055e-09 18-37 PR00011A 14.06 6.178e-09 18-37
139	BL00472	Small cytokines (intercrine/chemokine) C-C subfamily signatur.	BL00472C 20.76 6.850e-34 65-102 BL00472B 14.67 8.579e-15 44-62 BL00472A 7.45 2.895e-11 11-23
140	PR00988	URIDINE KINASE SIGNATURE	PR00988A 6.39 8.905e-09 2-20
141	BL00388	Proteasome A-type subunits proteins.	BL00388A 23.14 5.500e-39 8-54 BL00388B 31.38 1.310e-27 66-108 BL00388C 18.79 3.842e-10 121-143
142	PD02699	PROTEIN DNA-BINDING BINDING DNA.	PD02699C 24.84 1.000e-40 614-661 PD02699A 8.91 3.250e-35 235-264 PD02699B 18.28 6.571e-21 500-524
143	DM01970	0 kw ZK632.12 YDR313C ENDOSOMAL III.	DM01970B 8.60 2.612e-13 551-564
144	PR00624	HISTONE H5 SIGNATURE	PR00624G 4.08 6.900e-09 180-200
146	BL00038	Myc-type, 'helix-loop-helix' dimerization domain proteins.	BL00038B 16.97 9.027e-10 57-78
147	PR00209	ALPHA/BETA GLIADIN FAMILY SIGNATURE	PR00209B 4.88 4.494e-12 393-412
149	PR00179	LIPOCALIN SIGNATURE	PR00179B 9.56 2.895e-13 128-141 PR00179A 13.78 3.250e-11 40-53 PR00179C 19.02 6.040e-11 158-174
152	PR00929	AT-HOOK-LIKE DOMAIN SIGNATURE	PR00929C 5.26 8.759e-09 55-66
153	BL00352	'Cold-shock' DNA-binding domain proteins.	BL00352B 23.66 5.886e-25 119-158 BL00352A 12.19 1.429e-17 93-108
154	BL00352	'Cold-shock' DNA-binding domain proteins.	BL00352B 23.66 5.886e-25 119-158 BL00352A 12.19 1.429e-17 93-108
155	BL00299	Ubiquitin domain proteins.	BL00299 28.84 2.250e-25 32-84

SEQ ID NO:	ACCESSION NUMBER	DESCRIPTION	RESULTS
156	PF00534	Glycosyl transferases group 1.	PF00534B 14.47 8.364e-15 325-349
157	BL00028	Zinc finger, C2H2 type, domain proteins.	BL00028 16.07 8.615e-11 46-63
164	BL00880	Acyl-CoA-binding protein.	BL00880 17.52 8.875e-34 46-96
168	BL00740	MAM domain proteins.	BL00740B 19.76 3.813e-09 690-711
170	BL00368	Ribonucleotide reductase small subunit proteins.	BL00368A 36.98 1.000e-40 84-139 BL00368B 22.06 5.846e-26 161-187
172	BL01188	GNS1/SUR4 family proteins.	BL01188C 22.65 6.667e-10 154-205

TABLE 4

SEQ ID NO:	PFAM NAME	DESCRIPTION	p-value	PFAM SCORE
1	EMP24_GP25L	emp24/gp25L/p24 family	6.4e-90	312.2
2	mito_carr	Mitochondrial carrier proteins	1.1e-09	36.2
3	SCAN	SCAN domain	1e-56	201.9
4	ion_trans	Ion transport protein	5.5e-260	877.1
6	Skp1	Skp1 family	2.2e-64	227.3
12	zf-C2H2	Zinc finger, C2H2 type	3.2e-53	190.2
13	Rap_GAP	Rap/ran-GAP	5.4e-117	402.1
17	MSP_domain	MSP (Major sperm protein) domain	2.9e-10	47.5
18	pkinase	Eukaryotic protein kinase domain	0.022	12.5
21	DnaJ	DnaJ domain	2.4e-34	127.5
22	mito_carr	Mitochondrial carrier proteins	1.1e-85	296.7
24	PH	PH domain	2.4e-14	56.3
26	transport_prot	Serum albumin family	2.4e-06	11.0
28	zf-C2H2	Zinc finger, C2H2 type	2.7e-22	87.5
31	zf-A20	A20-like zinc finger	0.0056	23.3
32	Dihydrooratase	Dihydroorotase-like	7.6e-88	305.3
33	asp	Eukaryotic aspartyl protease	5.7e-211	714.3
34	SRCR	Scavenger receptor cysteine-rich domain	4.3e-187	634.9
37	F-box	F-box domain.	0.089	19.3
39	NADHdh	NADH dehydrogenases	1.3e-39	145.0
42	Ribosomal_L30	Ribosomal protein L30p/L7e	1.2e-10	40.0
43	G_Adapt_CT	Gamma-adaptin, C-terminus	1.7e-67	237.6
50	KRAB	KRAB box	7.4e-12	52.8
51	SCAN	SCAN domain	5.4e-67	236.0
55	PH	PH domain	1.1e-06	28.9
56	SH3	SH3 domain	1.8e-14	61.5
57	globin	Globin	2.1e-26	96.5
60	NHL	NHL repeat	1.7e-46	167.9
61	ATP-synt_A	ATP synthase A chain	1.3e-10	48.7
62	HSP90	Hsp90 protein	0	1548.4
65	annexin	Annexin	1.9e-77	270.7
66	ferritin	Ferritins	4.1e-114	386.1
67	PUA	PUA domain	1.7e-16	68.2
69	DUF34	Domain of unknown function DUF34	1.2e-33	125.2
73	ig	Immunoglobulin domain	0.0015	16.5
75	Exonuclease	Exonuclease	0.076	-29.3
76	globin	Globin	5.9e-53	189.1
78	globin	Globin	4.6e-57	203.0
79	K_tetra	K+ channel tetramerisation domain	5.6e-22	86.4
81	PH	PH domain	2.2e-21	81.4
82	Band_41	FERM domain (Band 4.1 family)	1.1e-22	73.0
85	oxidored_q1	NADH-Ubiquinone/plastoquinone (complex I)	3.6e-06	25.3
87	ECH	Enoyl-CoA hydratase/isomerase family	4.1e-52	186.6
88	ECH	Enoyl-CoA hydratase/isomerase family	4.1e-52	186.6
89	ECH	Enoyl-CoA hydratase/isomerase family	1.3e-20	81.9
91	Ribosomal_L6	Ribosomal protein L6	9e-99	341.6
95	Ribosomal_S6e	Ribosomal protein S6e	1.8e-90	314.0
96	arf	ADP-ribosylation factor family	5.4e-98	339.0
97	FG-GAP	FG-GAP repeat	1.4e-77	271.2

SEQ ID NO:	PFAM NAME	DESCRIPTION	p-value	PFAM SCORE
100	tubulin	Tubulin/FtsZ family	1.8e-49	177.8
101	tubulin	Tubulin/FtsZ family	1.8e-49	177.8
102	carb_anhydrase	Eukaryotic-type carbonic anhydrase	2.3e-48	174.1
104	GrpE	GrpE	5.9e-70	245.8
111	FAD_binding	FAD binding domain	7.2e-122	418.3
112	rrm	RNA recognition motif.	9.7e-12	52.4
113	SET	SET domain	3.3e-34	127.1
114	SET	SET domain	3.3e-34	127.1
115	fn3	Fibronectin type III domain	2.6e-292	984.5
126	oxidored_q2	NADH-ubiquinone/plastoquinone oxidoreduct	1.8e-44	161.2
132	Phosducin	Phosducin	0.097	9.7
139	IL8	Small cytokines (intecrine/chemokine), inter	3.6e-38	131.6
141	proteasome	Proteasome A-type and B-type	6.1e-39	142.8
142	RFX_DNA_binding	RFX DNA-binding domain	4.1e-43	156.7
143	RhoGEF	RhoGEF domain	8.1e-64	225.4
146	HLH	Helix-loop-helix DNA-binding domain	1.4e-05	31.9
147	BRCT	BRCA1 C Terminus (BRCT) domain	1.2e-28	108.6
149	lipocalin	Lipocalin / cytosolic fatty-acid binding pr	2.5e-41	146.4
150	Armadillo_seg	Armadillo/beta-catenin-like repeats	7.8e-10	46.1
153	CSD	'Cold-shock' DNA-binding domain	1.1e-33	124.6
154	CSD	'Cold-shock' DNA-binding domain	1.1e-33	124.6
156	Glycos_transf_1	Glycosyl transferases group 1	6.7e-39	142.7
157	zf-C2H2	Zinc finger, C2H2 type	0.00041	27.1
162	ig	Immunoglobulin domain	1e-06	26.7
164	ACBP	Acyl CoA binding protein	1.1e-39	145.2
165	DUF60	Domain of unknown function DUF60	2.2e-36	134.3
168	Gelsolin	Gelsolin repeat.	9.3e-109	374.7
169	LRR	Leucine Rich Repeat	0.00025	27.8
170	ribonuc_red	Ribonucleotide reductases	1.5e-122	358.6
172	GNS1_SUR4	GNS1/SUR4 family	2.7e-09	-26.6

TABLE 5

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number_corresponding SEQ ID NO: in priority application	SEQ ID NO: in USSN 09/515, 126
1	173	345	517	788CIP2_3	788_9798
2	174	346	518	788CIP2_5	788_10429
3	175	347	519	788CIP2_6	788_10672
4	176	348	520	788CIP2_7	788_10828
5	177	349	521	788CIP2_8	788_10879
6	178	350	522	788CIP2_9	788_11026
7	179	351	523	788CIP2_10	788_11045
8	180	352	524	788CIP2_11	788_11381
9	181	353	525	788CIP2_12	788_11481
10	182	354	526	788CIP2_13	788_11927
11	183	355	527	788CIP2_14	788_12700
12	184	356	528	788CIP2_15	788_12706
13	185	357	529	788CIP2_16	788_12741
14	186	358	530	788CIP2_17	788_12744
15	187	359	531	788CIP2_18	788_13017
16	188	360	532	788CIP2_20	788_13633
17	189	361	533	788CIP2_21	788_13637
18	190	362	534	788CIP2_22	788_13645
19	191	363	535	788CIP2_23	788_13695
20	192	364	536	788CIP2_24	788_13756
21	193	365	537	788CIP2_25	788_13858
22	194	366	538	788CIP2_26	788_13886
23	195	367	539	788CIP2_27	788_14039
24	196	368	540	788CIP2B_2	788_31
25	197	369	541	788CIP2B_3	788_54
26	198	370	542	788CIP2B_4	788_58
27	199	371	543	788CIP2B_5	788_768
28	200	372	544	788CIP2B_6	788_1718
29	201	373	545	788CIP2B_7	788_1748
30	202	374	546	788CIP2B_8	788_1751
31	203	375	547	788CIP2B_9	788_1850
32	204	376	548	788CIP2B_10	788_1946
33	205	377	549	788CIP2B_11	788_1967
34	206	378	550	788CIP2B_12	788_1994
35	207	379	551	788CIP2B_13	788_2331
36	208	380	552	788CIP2B_14	788_2725
37	209	381	553	788CIP2B_15	788_2735
38	210	382	554	788CIP2B_16	788_2961
39	211	383	555	788CIP2B_18	788_3039
40	212	384	556	788CIP2B_19	788_3287
41	213	385	557	788CIP2B_20	788_4305
42	214	386	558	788CIP2B_21	788_4356
43	215	387	559	788CIP2B_22	788_4389
44	216	388	560	788CIP2B_24	788_4458
45	217	389	561	788CIP2B_25	788_4685
46	218	390	562	788CIP2B_26	788_4825
47	219	391	563	788CIP2B_27	788_4829
48	220	392	564	788CIP2B_28	788_4856

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number_corresponding SEQ ID NO: in priority application	SEQ ID NO: in USSN 09/515, 126
49	221	393	565	788CIP2B_29	788_4917
50	222	394	566	788CIP2B_30	788_4939
51	223	395	567	788CIP2B_31	788_4943
52	224	396	568	788CIP2B_32	788_5175
53	225	397	569	788CIP2B_33	788_5254
54	226	398	570	788CIP2B_34	788_5566
55	227	399	571	788CIP2B_35	788_5725
56	228	400	572	788CIP2B_36	788_5794
57	229	401	573	788CIP2B_37	788_5796
58	230	402	574	788CIP2B_38	788_5841
59	231	403	575	788CIP2B_39	788_5922
60	232	404	576	788CIP2B_40	788_5942
61	233	405	577	788CIP2B_41	788_5980
62	234	406	578	788CIP2B_42	788_6025
63	235	407	579	788CIP2B_43	788_6151
64	236	408	580	788CIP2B_44	788_6849
65	237	409	581	788CIP2B_45	788_7191
66	238	410	582	788CIP2B_46	788_7281
67	239	411	583	788CIP2B_47	788_8402
68	240	412	584	788CIP2B_48	788_9636
69	241	413	585	788CIP2B_49	788_9645
70	242	414	586	788CIP2B_50	788_9784
71	243	415	587	788CIP2B_51	788_10075
72	244	416	588	788CIP2B_52	788_10223
73	245	417	589	788CIP2B_53	788_10369
74	246	418	590	788CIP2B_54	788_10413
75	247	419	591	788CIP2B_55	788_10441
76	248	420	592	788CIP2B_56	788_10607
77	249	421	593	788CIP2B_57	788_10625
78	250	422	594	788CIP2B_58	788_10857
79	251	423	595	788CIP2B_59	788_10868
80	252	424	596	788CIP2B_60	788_10962
81	253	425	597	788CIP2B_61	788_11218
82	254	426	598	788CIP2B_62	788_11224
83	255	427	599	788CIP2B_63	788_11251
84	256	428	600	788CIP2B_64	788_11262
85	257	429	601	788CIP2B_65	788_11597
86	258	430	602	788CIP2B_66	788_11708
87	259	431	603	788CIP2B_67	788_11719
88	260	432	604	788CIP2B_68	788_11719
89	261	433	605	788CIP2B_69	788_11719
90	262	434	606	788CIP2B_70	788_11823
91	263	435	607	788CIP2B_71	788_11865
92	264	436	608	788CIP2B_72	788_11939
93	265	437	609	788CIP2B_73	788_11961
94	266	438	610	788CIP2B_74	788_12119
95	267	439	611	788CIP2B_75	788_12128
96	268	440	612	788CIP2B_76	788_12190
97	269	441	613	788CIP2B_77	788_12212
98	270	442	614	788CIP2B_78	788_12233
99	271	443	615	788CIP2B_79	788_12392

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number_corresponding SEQ ID NO: in priority application	SEQ ID NO: in USSN 09/515, 126
100	272	444	616	788CIP2B_80	788_12518
101	273	445	617	788CIP2B_81	788_12518
102	274	446	618	788CIP2B_82	788_12525
103	275	447	619	788CIP2B_83	788_12594
104	276	448	620	788CIP2B_84	788_12737
105	277	449	621	788CIP2B_85	788_12818
106	278	450	622	788CIP2B_86	788_12825
107	279	451	623	788CIP2B_87	788_12868
108	280	452	624	788CIP2B_88	788_13029
109	281	453	625	788CIP2B_89	788_13031
110	282	454	626	788CIP2B_90	788_13037
111	283	455	627	788CIP2B_91	788_13051
112	284	456	628	788CIP2B_92	788_13061
113	285	457	629	788CIP2B_93	788_13105
114	286	458	630	788CIP2B_94	788_13105
115	287	459	631	788CIP2B_95	788_13131
116	288	460	632	788CIP2B_96	788_13281
117	289	461	633	788CIP2B_97	788_13325
118	290	462	634	788CIP2B_98	788_13453
119	291	463	635	788CIP2B_99	788_13557
120	292	464	636	788CIP2B_100	788_13638
121	293	465	637	788CIP2B_101	788_13638
122	294	466	638	788CIP2B_102	788_13655
123	295	467	639	788CIP2B_103	788_13656
124	296	468	640	788CIP2B_104	788_13668
125	297	469	641	788CIP2B_105	788_13681
126	298	470	642	788CIP2B_106	788_13725
127	299	471	643	788CIP2B_107	788_13796
128	300	472	644	788CIP2B_108	788_13802
129	301	473	645	788CIP2B_109	788_13814
130	302	474	646	788CIP2B_110	788_13816
131	303	475	647	788CIP2B_111	788_13836
132	304	476	648	788CIP2B_112	788_13852
133	305	477	649	788CIP2B_113	788_13855
134	306	478	650	788CIP2B_114	788_13990
135	307	479	651	788CIP2B_115	788_14056
136	308	480	652	788CIP2B_116	788_14066
137	309	481	653	788CIP2B_117	788_14070
138	310	482	654	788CIP2C_1	788_360
139	311	483	655	788CIP2C_2	788_1175
140	312	484	656	788CIP2C_3	788_3964
141	313	485	657	788CIP2C_4	788_5522
142	314	486	658	788CIP2C_5	788_5537
143	315	487	659	788CIP2C_6	788_5573
144	316	488	660	788CIP2C_7	788_5868
145	317	489	661	788CIP2C_8	788_5868
146	318	490	662	788CIP2C_9	788_6265
147	319	491	663	788CIP2C_10	788_7013
148	320	492	664	788CIP2C_11	788_7017
149	321	493	665	788CIP2C_12	788_7019
150	322	494	666	788CIP2C_13	788_8583

SEQ ID NO: of full-length nucleotide sequence	SEQ ID NO: of full-length peptide sequence	SEQ ID NO: of contig nucleotide sequence	SEQ ID NO: of contig peptide sequence	Priority docket number_corresponding SEQ ID NO: in priority application	SEQ ID NO: in USSN 09/515,126
151	323	495	667	788CIP2C_14	788_8792
152	324	496	668	788CIP2C_15	788_8816
153	325	497	669	788CIP2C_16	788_8916
154	326	498	670	788CIP2C_17	788_8916
155	327	499	671	788CIP2C_18	788_8950
156	328	500	672	788CIP2C_19	788_9067
157	329	501	673	788CIP2C_20	788_9254
158	330	502	674	788CIP2C_21	788_9285
159	331	503	675	788CIP2C_22	788_10049
160	332	504	676	788CIP2C_23	788_10858
161	333	505	677	788CIP2C_24	788_11054
162	334	506	678	788CIP2C_25	788_11872
163	335	507	679	788CIP2C_26	788_12230
164	336	508	680	788CIP2C_27	788_12369
165	337	509	681	788CIP2C_28	788_12464
166	338	510	682	788CIP2C_29	788_12708
167	339	511	683	788CIP2C_30	788_13024
168	340	512	684	788CIP2C_31	788_13199
169	341	513	685	788CIP2C_32	788_13601
170	342	514	686	788CIP2C_33	788_13666
171	343	515	687	788CIP2C_34	788_13749
172	344	516	688	788CIP2C_35	788_14042

TABLE 6

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
517	A	40	359	LKIPMQFLHSGFWFSFFVFFGF*KFGFGPQGGRQGG\NKTKGEKLPPGKKKIPGPNPQENREKKGPPKTLKKFGNLKKKGGKTRGPRGEKNSDPKGTPGQNPGNKKK
518	A	50	371	VVVAPVVVGSGRA/PRHPAPAAMHSRRPDGFDGLGYRGGARDEQGFGGAFPARSFITGSDLGHWVTTLTDI\PGRRN\LHWGVKSPPYGVPTNCTPYEGPTEDPFDSGSG
519	A	114	380	DGFPLSFYNTPKSPTSEILFPLSEILFSGPFGWLAAPSCFGFSCHFLKKGFLGPGTVAHSCNPSTLGGRGGWI/T/RGQEIKTILANTVKP
520	A	13	411	GPLRVDTWSNVESLACAQREEEEENERKKLPGTASPEKKQE\LVDEPAVGESKEEKIELKSITADGESPPASKINMDALQPNENEDKSPYPTPESTGEEDEEEPEMPVGPRPRPLSELHLKEKTVPMPEASAC
521	A	1	451	KINTELQTEVAMLKSMVLWLGEQVQSLQLQQQLRHHFNHIHICVTNSEYNQSEYPWDLVKAHLQGAFTSNITFDTGEFQNTIIDSTK*PHDL*ASLDDCS*I*DGLSILASVLMLAAPVLALQTPSWMLFMALPVSLSRLKAYFLVSVLS
522	A	354	549	IFWVGKNALHEDE*NTERRTDDIPVWDHEFLKVDQGTLFELILAANYLYIRGWLDVTCKTGANMI
523	A	448	489	YARSAIGLLVFQSLMLVAAGDQFLFRP*GE*RQPR**LKMTKHPPNRRGISFEVGAQLEARDRLKNWYPAHIEDIDYEEGKVLIHFKRWNHRYDEWFCWDSPYLRPLEKIQLRKEGLHEEDGSSVSNKSGKFCSMQGQPLGCLCFNL
524	A	35	402	KLFLGFPILHGSSSPP*LKVPSLTHKVSALVSSNSF*VMYLVIFHMNMKVAWRRGSLLHSFQLGFYKSQHFGRPRWADRLRLGVRDQPGQHGETPSLLVQKKI
525	A	203	3	GPKPFSLGHKGKSCVKKNFFFFFGGSHL*MQLLRRLRWEDQLSSGGRGCSGP*LHHCTVAWATEQD
526	A	617	14	VFTGNYNSKMRIWVGTQPNHISAHPHWVMVISVSSRNALIDMPRNPCLTSCVCLSIQSSRRLRLTIRIFMPDSWLKSCLTSYVCLSIQSSRCLRLTIRIFMPDSWLKSCLTSYVCLSIQSSRRLQLTIRIFMPDSWLKSCLTSYVCLSIQSSRCLQLTIRIFMPDSWLKSCLTSYVCLLIQSSRCLKLTVRIFMPLVPRLA
527	A	149	759	SRMTKKRKRQHDFQKVKLKVGKKKPK

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				LQNATPTNFKTKTIHLPEQLKEDGTLPTNNRKLNIKDLLSQMHHYNAGVKQRALLGLKDLLSQYPFIIDAHLLNILSEVTAVFTDKDANVRLAAVQLLQFLAPKIRAEQISPFFPLVSAHLSSAMTHITEGIQEDSLKVLDILLEQYPALITGRSSILLKNFVELISHQQLSKGLDK
528	A	117	369	WLPGEMAVEPPADRLQEPLTFRDGAEDFTQEEWGQLDPTQRILYRDVMLETFDHLLSIGPELTKPEDISQLEQGTDLWVAETLT
529	A	836	433	AVGGSSRSHGYRRPAGTRPVSHAQDPSAPPPAS/RLPCIEARGGQ*SLPSLILQPGGAGWGILLPNSVLPPAPCTPQSHVSLPSAGPCAPLGGRRQAPLGPRAAEPCRVRGCGAAALSCALPALGRGTFGWKTGI
530	A	1282	210	VRCTLSPACQTPPQGPHRSSGPVLSSPCPAPVGEWLPPNGP*SDDELPLSQQCKCRTKSLAPPSSA*GS*HRT*LPGTGH*CGLARIESSAGVLEEASAAK*GSGEAPGSKQHSSSPAATSAPSAGSWHPGTAAAHPPG/SSASFHTFAAFTVSLSASELGS**QSRVKHFFLSCSSTSCCSRALIHTKSSESGMSSRRSAFSSPSSGGISLSWLGPKSRISSVTSSRCSRVS*NFLSSEVTFVS\VSSHSSPQATLE*LCVKSQTSTMLSYAVELPSMPEPAPLLG*GLSHSHPHAVRDSGEAC*SMGSVTGPLYSGYKEEVVCCTLVEVFPSLLQVSRNPRMPFDFLGILICTPWGS
531	A	12	245	CAKNECQGGQVESGSF*AIYL*TVI\YLCVYLSVCLSIYLSVYLSIYLSIYLSIYLSIYPSIHPSIQLASI*SIFLWG
532	A	1	849	RKMAGSPELVVLDPPWDKELAAGTESQALVSATPREDFRVRCTSKRAVTEMLQLCGRFVQKLGDALPEEIREPALRDAQWTFESAVQENISINGQAWQEASDNCFMDSDIKVLEDQFDEIIVDIATKRKQYPRKILECVIKTIKAKQEILKQYHPVVHPLDLKYDPDPAPHMENLKCRGETVAKEISEAMKSLPALIEQGEGFSQVLRMQPVIHLQRIHQEVFSSCHRKPDAKPENFITQIETTPTETASRKTSDVVLKRKQTKDCPQRKWYPLRPKKINLDT
533	A	120	804	DCKNRNCSKGDSVPMHQQKRQPELVEGNLPVFVFPTELIFYADDQSTHKQVLTLYNPYEFALKFKVLCTTPNKYVVVDAAGAVKPQCCVDIVIRHRDVRSCHYGVIDKFRLQVSEQSQRKALGRKEVVATLLPSAKEQQKEEEEKRLKEHLTESLFFEQSFQPENRAVSSGPSLLTVFLGVVCIAALMLPTLGDV

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				ESLVPLYLHLSVNQKLVAAYILGLITMAI LRT
534	A	479	1906	SMISFCPDCGKSIQAAFKFCPYCGNSLPV EEHVGSQTFVNPHVSSFQGSKRGLNSSF ETSPKKVKWSSTVTSPRLSLFSDGDSSES EDTLSSSERSKGSGSRPPTPKSSPQKTRK SPQVTRGSPQKTSCSPQKTRQSPQTLKRS RVTTSLEALPTGTVLTDKSGRQWKLKSF QTRDNQGILYEAAPTSTLTCDSGPQKQK FSLKLDAKDGRLFNEQNFFQRAAKPLQV NKWKKLYSTPLLAIPTCMGFGVHQDKY RFLVLPSLGRSLQSALDVSPKHVLSERSV LQVACRLLDALEFLHENEYVHGNVTAE NIFVDPEDQSQVTLAGYGFAFRYCPSGK HVAYVEGSRSPHEGDLEFISMDLHKGCG PSRRSDLQSLGYCMLKWLYGFLPWTNC LPNTEDIMKQKQKFVDKPGPFVGPCGH WIRPSETLQKYLKVVMALTYEEKPPYA MLRNNLEALLQDLRVSPYDPIGLPMVP
535	A	3828	2909	AECEEVRRKSELFNPVSLDCKLRQKAIA EVDVGTDKAQNSDPILDTSSLVPGCSSV DNIKSSQTSQNQGLGRPTLEGDEETSEVE YTVNKGPASSNRDRVPPSSEASEHHPRH RVSSQAEDTSSSFDNLFIDRLQRITIADQ GEQQSEENASTKNLTGLSSGTEKKPHYM EVLEMRAKNPVPQLRKFKTNVLPFRQN DSSSHCQKSGSPISSEERRRRDKQHLDDI TAARLLPLHHMPTQLLSIEESLALQKQQ KQNYEEMQAKLAAQKLAERLNIKMRSY NPEGESSGRYREVRDEDDDWSSDEF
536	A	45	257	TLEFESHIIFMSHRIFFFCYLKAKIKILGW ARWLVPVIPALWEAEVGRSFEARSSRTA WATYGGSLFYKK
537	A	1071	472	KMACNIPNQRQRTLSTTGEALYEILGLH KGASNEEIKKTYRKLALKHHPDKNPDDP AATEKFKEITDAHAILTDISKRSIYDKYG SLGLYVAEQFGDENVNTYFMLSSWWA KALFVIVGLLTGCYFCCCLCCCCNCCCG HCRPESSVPEEDFYVSPEDLEEQIKSDME KDVDFPVFLQPTNANEKTQLIKEGSRSY CTDS
538	A	1	1528	RLAAAGADPAGSRGGRGSREAPTEAAR CRASPPPRAGAMRGSPGDAERRQRWGR LFEELDSNKDGRVDVHELRQGLARLGG GNPDPGAQQGISSEGDADPDGGLDLEEF SRYLQEREQRLLLMFHSLDRNQDGHIDV SEIQQSFRALGISISLEQAEKILHSMDRDG TMTIDWQEWRDHFLLHSLENVEDVLYF WKHSTVLDIGECLTVPDEFSKQEKLTGM

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				WWKQLVAGAVAGAVSRTGTAPLDRLK VFMQVHASKTNRLNILGGLRSMVLEGGI RSLWRGNGINVLKIAPESAIKFMAYEQIK RAILGQQETLHVQERFVAGSLAGATAQT IIYPMEVLKTRLTLRRTGQYKGLLDCAR RILEREGPRAFYRGYLPNVLGIIPYAGIDL AVYETLKNWWLQQYSHDSADPGILVLL ACGTISSTCGQIASYPLALVRTRMQAQA SIEGGPQLSMLGLLRHILSQEGMRGLYR GIAPNFMKVIPAVSISYVVYENMKQALG VTSRSFC
539	A	474	1098	RRRCFQGFLKKKKIIQPKGEGTGRGRGV VPFIHNIVNIHGLYYMCFSLVFACLSVRL SVWGLDLRSRAPLLPQPPPPCPTLGTELE QVWPPNHRWEPAGSAWATASTLAPESC VCKCVCVYACLSMCMYVHGEPVCKYV QRETVTVTEAGGRWGCWRRGLPGADSS QVDPPQVSGRCDRVALRAELLGTQGPQ PALLAVHVPTLDGR
540	A	1178	1	VSFLLWTSGNKEDGRNGPAVPSRHTGG SRTWRMPK*QEGENPCPLPKRNLGNA*T FNS*KRTGPG/SPDSKGPDLGPI/SVPHILS GSPGPSS\PTLPQLHDLLAIPVMQGREAG GLRGHRALPLA*EEAGRALQLPLSQPPP CRPAAPFFNQ/GPQGCCPWRGSGHSPGP QLGG*EPGAYSGWQQTSRMSGVAKPGS GVRPATGSPTRVPPGACSQSLPGTAPSPR PPGPPALLRPAGVGALQGHQSRWQRQT PRAREPRPVPLCSLPAPPGSVGPGPRDSQ RRIAGCP*GQTSIFGQCHWGSVGSLGPPR CAGSPCGPLGSR*VHSGLHSSRSASWLP REQQGS/PGGGQGPLVWPHSSALRPPG/V PGNLPPPIAPKPKRAIPARDPAPPPGLQH E
541	A	1	438	MLAIFKYLINNRLITT*Q/WTEKLTS*QMI TIHNTKGLT*SLILLSLIIFIATTNLLGLLP YSFTPTTQLSINLSMAIPL*TVAEFIGFRS KIKNALSHFLPQGTPTPLIPILVIIETISLLI QPILLDVLLTANITAGHLLIHLN
542	A	13	577	LNINKKYNYIKR*DYYKIII\NKKMNII*IIN K*YNKE*YK*NMKKIIK\Y*IYYSQVSTPT IVKVSRNLGKVGSKCCKHPEAKRMPCA EDYLYVVLNQL*VLHEKTPVNDRVTKC *KESLVNRRPCFSTLEVDKTYVTKKFNA ETFTFHADICTLSEKERQIKKQTALDELV KHKPKATKEQLKAVMDD
543	A	297	0	SSSSSSPSTPGPPSPFLRQPPFGLNPFLPAG D*GTQRGGSGRGGRGRGR/DGSTNEGRG ERRGQG

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
544	A	395	1	AKRKSVPEEIWKSRGQFKNQQLNKENNLGQEIATCTKIPSRKRDIEFNEFVKNFTVRSILVAEQIDPMEENCHKYGTC*KMLKQNSDLIIQ\ESMMEKKKPCKYSECGRTFRGHITLVQHQITHCGERPCI
545	A	291	209	NNKNKIASKRKKNSPEPGGLTTEL/YQTSKEELEPILLNLFQKIEKEEILSNSFRETSIALIPKPGKDITKEVNHRPISLMNRDAKVLNKMLSN*IQ
546	A	419	3	GKPPFYFGGPGFFPPPV/YFKPPPPKNFFGPPKKKKFPPPPGVYFFFFLRPPPPLFFFFFFFFFFFF*EGGGEKPSPP/IKKGFFKPKKFFSPPPGGKKKPPFLKKKKKKKDVPQSSNLESSLSFPIHKTGTIIPASLNV
547	A	28	414	HTLLLQCLSAVAFSQPCPHPSRL/CNPCCASGPPCSS*ACCPRFCQSCPVC*PQNLLSSCPSPDLALTSPSAGRPRCWWNCCQLTPCLSLQTQEWTGAHLQLRRWGTQAGPEEPVHQRTVGSTSHSCSYA
548	A	419	2	RWSARSRITGEEGSESAVQER*LTYHTHGGIANRDRTHCPVSLVN\VSSMSAGDVITAGKMQGKVVLAEPTAALCHGN\GLHYYRQDWSHAAAYVTVPPVRVDTNTSTYLMSLLANDTLNIVASDRRPFTTKQKAMGNV
549	A	201	550	GCLVSLGVSHAWQRVGTWEIFVG*MNAG*MSA*TRGTEISRIGQWMGGEGMSGLTSWFLLGEMNP*GLRELNLLLTLSITMNGETIACAEGCQAICDTGTSLLT/GPTSPIAQHQSDI
550	A	1	526	LLDNCQMTVLGAPTCIHGNTVSVICTGSLTQPLFPCLANVSDPYLSAVPEGSALICLEDKRLRLVDGDSRCAGRVEIYHDGFWGTICDDGWDLSDAHVVCQKLSCGVAFNATVSAHFG/VGAGPIW/LDDL/NCTGMESHLRQ/CPF
551	A	369	130	FFFFFFXXXGVXGLNFFFFXFFXXFFFFFFFXFXXXXXFFFFFXGXXXXFXXFXXXXXFFFFFFFFFFFFFFFFFFFFXCY
552	A	255	338	EVMHFCLFVF*DRVLLCHPGWSAVAQSQ
553	A	120	383	TAPKKKKGGAPLGGPPRGPILPRIPENRRPP**GALLKPSPGPRFNTAWPGKPAPRESLKRAPSLGGGPKRPNPPPEVKASGADKKL
554	A	3	365	YHIVQPSP*PLAGALSALLMTSGLAMRSHFHSITLLILGLLSSTLTIYQ*WRDVTRESTYQGHHTPPVRKGLRYGIILFITSEVIFFAGLF*AFYHSSLAPTPQLGGHWPPTGITPLNTLE

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
555	A	1	125	LMNSNSIPPIPLRPTHTPPMKKL*IRTAYPRFRYDQLIHLL*KNFLPLTLALLI*HVKKKKKK
556	A	372	86	PACGGFSPPPPPKFFFFPPPLFFFGGGPPFPPPPKGFFFPNPPRGFFFPPP*KKKFFFPPPPFFCPPPVFFFCPPPLFFFFFFFFFFFFFFFF
557	A	2	376	FVAAEVRDGGGGAGTRPPSGRMKRRNADCSKLRRPLKRNRITEGIYGSTFLYLKFLVVWALVLLADFVLEFRFEYLWPFWLFIRSVYDSFRYQGLVCLVIHVFHVWGPFSCLE*MLLAILR*LC
558	A	39	459	VLNHSYMIPGRRNVTRCISRQLEVQPHALELPDKCSLAFVVRIKRIDGGSLLVQRTIARLCLKKIFSGVFLKMLRIVEPYVT*GFPNLKSVWEFILKHGQAKVKNKTIGQTQWLTPVIPALWEAKDGWIMRSGDRDHPGQ
559	A	2	456	EVRSHVKVLGEMLVMYRRPGQAPPDQEALQVVYERCEKLRPTLFRLASDTTDDDDALSEILQANDLLTQGVLLYKQVMEGRVTFGNRVTSSLGDIPVSRVFQNPAGCMKTCPLIDLEVDNGPAQMGTVVPSLLHQDLAALGNSDAPVTGMV
560	A	54	1063	RRRADGCIYGVSRRARVVAYRRDEMWSEGRYEYERIPRERAPPRSHPSDESGYRWTRDDHSASRQPEYRDMRDGFRRKSFYSSHYARERSPYKRDNTFFRESPVGRKDSPHSRSGSSVSSRSYSPERSKSYSFHQSQHRKSVRPGASYKRQNEGNPERDKERPVQSLKTSRDTSPSSGSAVSSSKVLDKPSRLTEKELAEAASKWAAEKLEKSDESNLPEISEYEAGSTAPLFTDQPEEPESNTTHGIELFEDSQLTTRSKAIASKTKEIEQVYRQDCETFGMVVKMLIEKDPSLEKSIQFALRQNLHEIGERCVEELKHFIAEYDTSTQDFGEPF
561	A	1887	442	LLIREVFIQVLIKLRVCFVQCFSEADRDIMTLANHWNCPVLSSDSDFCIFDLKTGFCPLNSFQWRNMNTIKGTQNYIPAKCFSLDAFCHHFSNMNKALLPLFAGLCGNDHVNLPIMETFLSKARLPLGATSSKGRRHHRILGLLNWLSHFANPTEALDNVLKYLPKKDRENVKELLCCSMEEYQQSQVKLQDFFQCGTYVCPDALNLGLPEWVLVALAKGQLSPFISDALVLRRTILPTQVENMQQPNAHRISQPIRQIIYGLLLNASPHLDKTSWNALPPQPLAFSEVERINKNIRTSIIDAVELAKDHSDLSRLTELSLRRRQMLLLETLKVKQTILEPIPTSLKLPIAVSCYWLQHTETKAKLHHLQSLLLTMLVGPLIAIINSPGKEELQEDGA

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				KMLYAEFQRVKAQTRLGTRLDLDTAHIFCQWQSCLQMGDVSQPAAVHSSPRARPNSTVQWKPGARTMPATASIDLCRKSPEHMS
562	A	180	2	TRHFYLKKNNFPPTQKSCLPLFSLSSLCVCVCVCVCVCVCVCVYSSAFKKTRRMHRWIR
563	A	3	301	GLVNMALYVSPIVSGEVIRSRGGSTSEFTPGYVKPKHEVNPQMTLRRLPDEDPQNLADPAYRRRRIIMQNMRDEELAIAQVEEMQAVSAVLKGKYTMTG
564	A	270	31	SFVVEELSWALQMLGRIPGLFAQIQNLGKEKIFWKEFSILDANKNTDVSWKEIKIATLTGVWKKLILTFETVSYKSQNQH
565	A	341	3	YDVIVGGELFSDYSDRPRKFVALNPKLKSTGAGRYQFLSRWWDAYRKQLGLKDFSPKSQAAVALQQIKERGALPMIDRGDIRQAIDRCSNMWASLPGAGYGQFEHKADSLIA
566	A	2	265	QEEWSLLSEAQRCLYHDVMLENLTLISSLGCWYGAKDETPSKQTLSIQQESPLRTHWTGGNDVTLKLRLEYTGMVIPHCGLKLLGLK
567	A	376	206	TPGWTQLDSSQVNLYRDEKQENHSSLVSLGKTETLICLKISWHSLRIRTYWAVRSC
568	A	212	428	TQQQAARSSEAHSWQSSSTQSLLEKFFLWRYSRGLRISLWFETESHCVPQVGVRWCDLGSLKPLPPGFRRFS
569	A	3	764	RDKHINNLKKKCQKESEQNREKQQRIETLERYLADLPTLEDHQKQSQQLKDSELKSTELQEKVTELESLLEETQAICREKEIQLESLRQRETEFSSAGHSLQDKQSVEETSGEGPEVEMESWQKRYDSLQKIVEKQQQKMDQLRSQVQSLEQEVAQEEGTSQALREEAQRRDSALQQLRTAVKELSVQNQDLIEKNLTLQEHLRQAQPGSPPSPDTAQLALELHQELASCLQDLQAVCSIVTQRAQGHDPNLSL
570	A	110	498	KVFCQMNSLLDARSLSLTSMLFVCEVCVKMCDPAKGAAGQRTIAALLPCLLDKGMMSTVTEVRALSINTLVKISKSAGAMLKPHAPKLIPALLESLSVLEPQVLNYLSLRATEQEKAAMDSARLSAAKS
571	A	268	464	ALTSPTLPQPLPGATVIEPLDDISAVTDILTHREGARLETPPPWLAMFTDQPALPNPCSPASVGP
572	A	198	878	GPRQKRKPKSSSSGTKGKGNLEPYLNFLPLSQVNNKFGLRCKNCKTNIHEHCQSYVEMQRCFGKIPPGFHRAYSSPLYSNQQYA

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				CVKDLSAANRNDPVFETLRTGVIMANK ERKKGQADKKNPVAAMMEEEPESARPE EGKPQDGNPEGDKKAEKKTPDDKHKQP GFQQSHYFVALYRFKALEKDDLDFPPGE KITVIDDSNEEWGRGKIGEKVGFFPSNFII WGPA
573	A	120	1	FRPADKTNVKAAWGKVGAHAGEYGAE ALERMFLSFPTTCI
574	A	384	1	RTDSSTTNRESVNSEDGDAKEPPYRGPF CGRDRVHTDFTPSPYDTDSLKLKKGDII DIISKPPMGTWMGLLNNKVGTFKFIYVD VLSEDEEKPKRPTRRRRKGRPPQPKSVE DLLDRINIKEHMPTCI
575	A	460	3	RVVQTVDPSKAQAKPHTLWVAFAKLYE DNGQLDDARVILEKATKVNFKQVDDLA SVWCQCGELELRHENYDEALRLLRKAT ALPARRAEYFDGSEPVQNRVYKSLKVW SMLADLEESLGTFQSTKAVYDRILDLRIA TPHVFGRSEDQASTR
576	A	396	2	CRECTEGEHAELPTVPLKDVVEQHKASL QVQLDAVNKRLPEIDSALQFISEIIHQLT NQKASIVDDIHSTFDELQKTLNVRKSVL LMELEVNYGLKHKVLQSQLDTLLQGQE SIKSCSNFTAQTLNHGTKA
577	A	1	45	LLYTYVLLASGVWVA*AYHRLIENNRN QIIQALLITILLGLYFTLLQASE\HLQSPFTI SDGIYGSTFFVATGFHRLHVIIGSTLLTIC FIPQLIFDFTSKHHFGIQAAV*YWHLVDG V*LFMHGSI*WVA
578	A	2091	320	ISHSLSTQDPNGESIIYNSSFSAVDSFAVS SPASDYLELDTIKNLVKKYSQFINFPIYV WSSKTETVEEPMEEEEAAKEEKEESDDE AAVEEEEEEKKPKTKKVEKTVWDWEL MNDIKPIWQRPSKEVEEDEYKAFYKSFS KESDDPMAYIHFTAEGEVTFKSILFVPTS APRGLFDEYGSKKSDYIKLYVRRVFITD DFHDMMPKYLNFVKGVVDSDDLPLNVS RETLQQHKLLKVIRKKLVRKTLDMIKKI ADDKYNDTFWKEFGTNIKLGVIEDHSNR TRLAKLLRFQSSHHPTDITSLDQYVERM KEKQDKIYFMAGSSRKEAESSPFVERLL KKGYEVIYLTEPVDEYCIQALPEFDGKR FQNVAKEGVKFDESEKTKESREAVEKEF EPLLNWMKDKALKDKIEKAVVSQRLTE SPWLLWVGQPSTDWSGNMRDFMKAQA YQPGK\DILPN\YYASQK\KTFEIN\PRPPL\ IRDMLRRIK\EDE\DDKTVLDLA\VVLV*N RQRLGSGYLFTQTLKAYGDW/RLERML RLSFEHLTPDAKGGRRAPKEEP*KRQQK

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				DTSRKTQSRDEE/MQEMGWWETDEEEE TSQRNPTA
579	A	25	378	IASGRPFFFFFFFFFFFFFFFFFFFFFFFF FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF FFFFFF*FYILKIFFFFFILEILCYLSNYGVG VTY/IRMSYYK*MLESYFFFTINIFFIGNN YLISLIYNIYAVKRQL
580	A	1	163	GSTTPAMEFASLFKKILLIDCRD/RGLAL LPRLVLSSWPQVIFLPWPPKFLGLRT
581	A	1	272	ESSCDWLNAYLTLVRCAQD*ADYFAER LYKSMKGAGTDEETLIRIIVTRAEVDLH GIKAKFQEKYQKSLSDMVRSDTFGDFRK LLVALLH
582	A	401	1	LYIFSPSRLIFFRGFSFFFFPPKKIFFLKIFP NDFFFPSHFFKIPSLGFFFYTFFIRKIFFFFP IIYFSPPGFFF*SPPFFFFFFFFFFFFFFFFF FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFG WYKQRPSQARVWTASLDFTL
583	A	764	139	GDLPDPCKFNFFPIPGPSLSSPPSPWIMFK KFDEKENVSNCIQLKTSVIKGIKNQLIEQ FPGIEPWLNQIMPKKDPVKIVRCHEHIEI LTVNGELLFFRQREGPFYPTLRLLHKYPF ILPHQQVDKGAIKFVLSGANIMCPGLTSP GAKLYPAAVDTIVAIMAEGKQHALCVG VMKMSAEDIEKVNKGIGIENIHYLNDGL WHMKTYK
584	A	1	343	FFFFEDKVSLCHPRWECSDDITAH\CSLN LGFK*SSHL\SLPK*LCLLGTQHYAPANF LYFVFFVETGFRHVGQAGLEHLGSTDPP ASA/FPKCWDYRCEPPCSTRFYYLS*LNF G
585	A	34	475	IPPTMGNLALGGKNSSRGFVDSLICNSSR AFMDWRALLSSLNDFASLSFAESWDNV GLLVEPSPPHTVNTLFLTNDLTDEVMDE VLLKKAHLILSYHPSIFRPMKRITWNTW KELLVIRALENRVGIYSPYTT*EA\APQG VNNWVA
586	A	17	1040	DGVSLLLYRLEYNGVISPHCNPGSSDSPA SASRVAGITGMRHHAQLIFVFLVETGFH HVGQAGLELLTSGDPPASASQSAGITGM SHHAWPDVNYLKNFQVILMGQG\WNPC SGPIRAPGCHKRMHALTPVGLGGFGNVL LLF*DG/CLSVAHPGAQWHDHSSLQPRP PGLR\YPPASAS*VSASGFHDDLGPRHTT LVLFRGQSCTRSPSATAPPRSTDMQS*A ALFPRGSCVGFLGPSVD/YRTPAEDLLG MPHCLCFPSSLNTQSGLAMMPLCLLASP PLEQPSWHATLLAAGHCAFCLYRSAVM LPK*QQQQQQQRLGTWLMPVFPAVRDQ

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				PGQHGETLSL
587	A	236	3	VDHLRSEV*DQPGQHIVKPRLY*KIQKLAGHGGACL*SPLLGKLRQENRLNPG\ARSCSEPRSHYCTPAWGTE*NSVS
588	A	93	413	GYCFSRRSGSMALWRAYQRALAAHPWKVQVLTADFWGGRQHCWPHASSYTCRPRPPGLKD/IASCSSLQEATSCPSAFCA*NSKATKQLVERRGLQEHQRGRTLTMVS
589	A	518	2	VGSQGLVPKKNRPAGKDLGAPSGGPPRKCIP/WQGLLLTAS\LLAL*EAPTTAWLFIASAPYEVAEGENVHLSVVYLRENLYSYGWYKGKTVEPNQLIAAYVIDTHVRTPGPAYSGRETISPSGDLHFQNVTLEDTGYYNLQVTYRNSQIEQASHHLRVYESVAQPSIQASSCI
590	A	121	473	VTRAHCWKMERSLRQSLLEPVPPSLPNSVLGWC/AELRVLCKLEENVQATNSPSEAEEFQLEVSGLLGEMNCPYLSLTSGDVTKRLLIQKNCLLLLTYLISELEAARMLCVNAPPKKS
591	A	243	353	DRVSLLLPRLDCNGTILVHCGLR/LPGFKRFS\CLSLPS
592	A	2	367	QPQTDTMVHLTPEEKAAATVLWGKGNVDEVGGEALVKLIGAYPWTQRFFESFGDLSTPDAVMGNPNGKAHGAKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDAEN/FMLLGNVLVCELA
593	A	158	418	IISVKRWMEPGQHGKTPPPLKIQRLAGH/SQVLRGLRHKNCLNPGDGGCSEPRVFHCTPAWVTEEDPVSLSKKREMDGNNTTFTPSQL
594	A	1	418	PQTQREPTMGLSAAD*TNV*AGWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGYAQDKGHGMKVADALTNAVAHVDDMPNALYALSDLHAHKL*VDPVNFKLLNHCLVVTLACHLLAEVTPGVHAFLDKFPG
595	A	76	440	SYPETWALEDASLEQMDNGYWGYMMTDPVTLNVGGHLYTTILTTLTRYPDSMLGAMNGGDFPTA*DPQGNYFIDLDGPLFRYVLNFLRTSELTLPLDFKEFDLLLK*SDFYQIEPLIQCIN
596	A	83	341	SFRRPMASASTQPAALSAEQAKASIH*FLALSAGPAVCSPTPTRDASLGCFWRAVVLAEVIQAFSAPENAVRMDEARDNACNDMGK
597	A	102	1	EHRPGVVAHACNPSTSGGQGEWIT*GQEFKTSLG
598	A	180	3	KKKKKKKKRGEIGKFYSLLEELTARSWF

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				SCFFIF*DKQGLPWWLGISYKGSGQYDIQA
599	A	64	326	GAEVAPLSIYFTQVPVVHIYSVIMTRLVKLAARKQSPFFLAVFGIYHLLITIHF*RQKKKIYLSIYLSIYLSIYLSIYLNDLTEENC
600	A	364	219	NFCTTAS*GRGCSEPR*RHCTPAWVTDYDSISNKN*TGGDFIKNIMRNLQDAIIGSVYFHVERVNRLN
601	A	61	284	EDHHAGTLITALSSH*FFT*VGLEINMLAFIPVLTKKKKKKRGGPFKGTKFKARGGGRKNFFKGAPKLNSGAGV
602	A	3	397	VSEEKLLVRIPRVIGATYSLANERLRALEDIEREIGAILQNAGTVILELSKEKTNERLLDRQAAAFTASVQHVEAELSAQIRYLTQVATGQPHEGSSYSSRKDCQMALKRVDYARLKLSDVARTCEQMLEN
603	A	110	370	GTYALKIIGHKRSSEGVQRHHLKGITEILMNRPSARNALGNVFVSELLETLAQLREDRQVRVLLFRSGVKGVFCAGADLLDREHMS
604	A	110	370	GTYALKIIGHKRSSEGVQRHHLKGITEILMNRPSARNALGNVFVSELLETLAQLREDRQVRVLLFRSGVKGVFCAGADLLDREHMS
605	A	110	370	GTYALKIIGHKRSSEGVQRHHLKGITEILMNRPSARNALGNVFVSELLETLAQLREDRQVRVLLFRSGVKGVFCAGADLLDREHMS
606	A	42	385	CSSEDKEDSFYPRNSKKKCKDFSKHQVFCFVLRRESCSVARAGVQWRDLGSLQAPPPGFTPFSWKGEDAVSLNRSIAPQPERQSETLGSNRLRRWRKNYKARRGGHMPVLQVLS
607	A	1	268	RMKTILSNQTVDIPENVDITLKGRTGIVKGPRGTLRRDFNHINVELSLLGKKKKRLRVDKWRGNTKELSTDRTMCIHCHNMIMVVIHVW
608	A	101	3	KHTLGWVQWLMPAIPALWEAKVGRSLETHASA
609	A	3	299	FHFFRKPPESKKPSVPETEADGFVLLGDTTDEQRMTARGKTSDIEANQPLETNKENSSSVTVSDPEMENKAGQTLENSSLMAELLSDVPVTLAPHVLAG
610	A	283	3	LALFYPQIKAIIFFFYLINASNGSILFFSGPFHTCSLKSSGALWCMIFRGRFHSLVCFSRLSGVLMEVEEHEVLQDSLDRCYSTPPMFFDIPD
611	A	2	363	NSAPEAGSCWKMKVNISFPATGCQKLIEVDDERKLRTFYDNRMAPEVAVDALVEE

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				WKGYVVRIIGGNDKQGFPMKHGDLTHGRVRLLLSKGHSCYRPRRTGERKRKSLRGCIVDANLSVL
612	A	56	415	ALIMSFIFEGIYDGFSSVLQFLGLYKKSGKLVFLGLDNAGKTTLLHMLKDDRLGQHVPTLHPTSEELTIAGMTFTTFDLGGHEQARRVWKNYLPAINGIVFLVDCADHSRLVESKVELNA
613	A	32	240	RGIPRLERPNPSLTLAPDTQDLFKLCCKCNKSDPVPPLTRKKKKKKKKKKKKKKKKKKKKKKKKRGAL
614	A	218	429	ELAPLQSLRKGPGLLSPPRACPVPTPAVSRTLLGNFEESLLRGRFAPSGHIDGYTAEIGASGSYCPQHVT
615	A	67	1204	VPAQPPKSRLSRGELLGLKRPVLVRPRWVGSEQAEATMEQCACVERELDKVLQKFLTYGQHCERSLEELLHYVGQLRAELASAALQGTPLSATLSLVMSQCCRKIKDTVQKLASDHKDIHSSVSRVGKAIDRNFDSEICGVVSDAVWDAREQQQILQMAIVEHLYQQGMLSVAEELCQESTLNVDLDFKQPFLELNRILEALHEQDLGPALEWAVSHRQRLLELNSSLEFKLHRLHFIRLLAGGPAKQLEALSYARHFQPFARLHQREIQVMMGSLVYLRLGLEKSPYCHLLDSSHWAEICETFTRDACSLLGLSVESPLSVSFASGCVALPVLMNIKAVIEQRQCTGVWNHKDELPIEIELGMKCWYHSVFACPILY
616	A	220	627	VGPKFEPDRSAGRRLAVVFRGNTPSLWLSMSIVTVQLGQCGNQIGFEVFDALLSDSHSSQGLCSMRENEAYQASCKERFFSEEENGVPIARAVLVDMEPKVINQTLSKAAQSGQWKYGQHACFCQKQGSGNNWAYG
617	A	220	627	VGPKFEPDRSAGRRLAVVFRGNTPSLWLSMSIVTVQLGQCGNQIGFEVFDALLSDSHSSQGLCSMRENEAYQASCKERFFSEEENGVPIARAVLVDMEPKVINQTLSKAAQSGQWKYGQHACFCQKQGSGNNWAYG
618	A	244	459	CTWNENQRMEIVWEVLFLLQANFIVCISAQQNSPKIHEGWWAYKEVVQGSFVPVPSFWGLVNSAWNLCSVGK
619	A	258	387	LNYHQNINDFFETESHSVTRLECSGTILAQCNLCLLSPSDSPP
620	A	4	281	RHIVIGEGKLKGQPGASVAAPCVRLVPHSIPVLALSPGLLCSARS\TKQKNSCQQLKEDVD\QTKWGTDTPSMD*ILMEEVKLEEQLKEVVEEYKLALSDTEGLQQSIQKLVDEANMCSIQGFCMDSLEVADVLENGTQCVS*QKNSCQQLKEDVDPNQMGDRYSLY

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				GLDTHGRSQVRRATEGGCGRI
621	A	648	937	RKGQQQTCLS\LRVSWLYRKVTFSIFFFF LIL*K*RWGLTMVTQAGLKLLGTSNPPA FALPKCWDYRCKSLPPAQKLTFSKATTK KTPQICQSMKVS
622	A	3	816	HETMEAAPSRFMFLLFLLTCELAAEVAA EVEKSSDGPGAAQEPTWLTDVPAAMEFI AATEDLEIPAVPILHSMVQKFPGVSFGIS TDSEVLTHYNITGNTICLFRLVDNEQLDL EDEDIESIDATKLSRFIEINSLHMVTEYNP VTVIGLFNSVIQIHLLLINE/PRPPQKYEE NMHRYQKAAKLFQGKIPPLFLVGQL\DE RKMGKVISFFKLKESQLPAL\QFTRL*DD EWDTLPTAEVSVEHVQNFCDGFLSGKLL KENRESEGKTPKVEL
623	A	15	286	DRVSL/LSPRLECSGVVSAHCKLRLPGSR /RFSCLSLS\GSWDYRRPPPRPANFFVFLV ETGFHCGLDLLTS*SARLGLPKCWDYKR EPPHPA
624	A	2411	325	NSSWPAEPAASPWRPLWRALGATFPSGS QPAARTPAGPCIGGMAPPGFKHVSSMLA \LTIIAST\WALTPTHYLTKHDVERLKASL DRPFTNLESAFYS\IVGLSSLG\AQVP\DA KKACTYI\RSNLDP\SNV\DSLFYGWPRAS Q\ALSGM*RSLFSNE\TKDLAFGQLFS*GT SSVYPRSYHAS/VAALKWALGLPLASQE ALSALTARLSKEETVLATVQALQTASHL SQQADLRSIVEEIEDLVARLDELGGVYL Q\FEEGLETTAL\FVAATYKA/LMDH\VG TE\PSIKE\DQVIQLMN\AIF\SKKNFE\SLS EAFSV\ASAAAVLSHNRYHVPVVVVPEG SASDTHEQAILRLQVTNVLSQPLTQATV KLEHAKSVAS\RATVLQKTSFTP\VGIVFE LNFMN\VKFS\SGY\YDFLGRKLKGDNRY IS\NTVELRVQDPPTEVGITNVDLSTV\DK DQSIAP\QTTRVTYP\AKAKGTFIADSHQ NFALFFQL\VGVNTGAELTPHQTFVRLH NQKTGQ\EVVFVAEPDNKNVYKFELDTS ERKGLNLTSRSGTYTFY\LIIGDATLKNPI LWNVADVVIKFPEEEAPSTVLSQNLFTP KQEIQHLFREPEKRPPTVVSNTFTALILSP LLLLFALWIRIGANVSNFTFAPSTIIFHLG HAAMLGLMYVYWTQLNMFQTLKYLAI LGSVTFLAGNRMLAQQAVKRTAH
625	A	498	915	QGSKVFIKKCST/SSLTREMQISTQRSHC GTLI*ASEVRK/SG*HQCQR*LWEGDL\L HC**QCELMEPTWKMAWHQLLEKKTHS PL/DLPVPHQGLDMAVM/CYQGA/CARN LYSQILTIAPNWNQPTYPSGREWMLKNA

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				RIAIG
626	A	130	1081	SVSTTRSFSVDSSAKTAAMPVTVTRTTIT TTTTSSSGLGSPMIVGSPRALTQPLGLLR LLQLVSTCVAFSLVASVGAWTGSMGNW SMFTWCFCFSVTLIILIVEL\SGLQARFPL SWRNFPITFACYAALFCLSASIIYPTTYV QFLSHGRSRDHAIAATFFSCIACVAYATE VAWTRARPGEITGYMATVPGLLKVLET FVACIIFAFISDPNLYQHQPALEWCVAVY AICFILAAIAILLNLGECTNVLPIPFPSFLS GLA\FCLSSSMPPPLFSGPSTSSMRSMAA SLGAREM*AAAAAMPTTCVPGTADWL WPS
627	A	1	2059	GCQRFMINMGDSHVDTSSTVSEAVAEE VSLFSMTDMILFSLIVGLLTYWFLFRKK KEEVPEFTKIQTLTSSVRESSFVEKMKKT GRNIIVFYGSQTGTAEEFANRLSKDAHR YGMRGMSADPEEYDLADLSSLPEIDNAL VVFCMATYGEGDPTDNAQDFYDWLQE TDVDLSGVKFAVFGLGNKTYEHFNAMG KYVDKRLEQLGAQRIFELGLGDDDGNL EEDFITWREQFWPAVCEHFGVEATGEES SIRQYELVVHTDIDAAKVYMGEMGRLK SYENQKPPFDAKNPFLAAVTTNRKLNQ GTERHLMHLELDISDSKIRYESGDHV\AV YPANDSALVNQLGKILGADLDVVMSLN NLDEESNKKHPFPCPTSYRTALTYYLDIT NPPRTNVLYELAQYASEPSEQELLRKMA SSSGEGKELYLSWVVEARRHILAILQDC PSLRPPIDHLCELLPRLQARYYSIASSSKV HPNSVHICAVVVEYETKAGRINKGVATN WLRAKEPVGENGGRALVPMFVRKSQFR LPFKATTPVIMVGPGTGVAPFIGFIQERA WLRQQGKEVGETLLYYGCRRSDEDYLY REELAQFHRDGALTQLNVAFSREQSHK VYVQHLLKQDREHLWKLIEGGAHIYVC GDARNMARDVQNTFYDIVAELGAMEH AQAVDYIKKLMTKGRYSLDVWS
628	A	1	1770	MSRRLLPRAEKRRRRLEQRQQPDEQRR RSGAMVKMAAAGGGGGGGRYYGGGSE GGRAPKRLKTDNAGDQHGGGGGGGG AGAAGGGGGGENYDDPHKTPASPVVHI RGLIDGVVEADLVEALQEFGPISYVVVM PKKRQALVEFEDVLGACNAVNYAADN QIYIAGHPAFVNYSTSQKISRPGDSDDSR SVNSVLLFTILNPIYSITTDVLYTICNPCG PVQRIVIFRKNGVQAMVEFDSVQSAQRA KASLNGADIYSGCCTLKIEYAKPTRLNV FKNDQDTWDYTNPNLSGQGDPGSNPNK

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				RQRQPPLLGDHPAEYGGPHGGYHSHYH DEGYGPPPPHYEGRRMGPPVGGHRRGP SRYGPQYGHPPPP\PPPPEYGPHADSPVL MVYGLDQSKMNCDRVFNVFCLYGNVE KVKFMKSKPGAAMVEMADGYAVDRAI THLNNNFMFGQKLNVCVSKQPAIMPGQ SYGLEDGSCSYKDFSESRNNRFSTPEQA AKNRIQHPSNVLHFFNAPLEVTEENFFEI CDELGVKRPSSVKVFSGKSERSSSGLLE WESKSDALETLGFLNHYQMKNPNGPYP YTLKLCFSTAQHAS
629	A	1826	5	SNLPEANVMSIVIPLGVDTAETSYLEMA AGSEPESVEASPVVVEKSNSYPHQLYTS SSHHSHSYNGFAPMRTHNRVLVPPPNTP APPPP\SVLISKNEVGHIYPLPNFDETSQC LPTISTS\EDGSYGTDVT\RCICGFTHDDG YMICCDKCSVWQHIDCMGIDRQHIPDTY LCERCQPRNLDKERAVLLQRRKRENMS DGDTSATESGDEVPVELYTAFQHTPTSIT LTASRVSKVNDKRRKKSGEKEQHISKCK KGSAPEIDPSSDGSNFGWETKIKAWMDR YEEANNNQYSEGVQREAQRIALRLGNG NDKKEMNKSDLNTNNLLFKPPVESHIQK NK\KILKSAK\DWPP\DALIIEYRGKFM\L R\EQFEANGYFF\KRP\YPFVLF\YSKF\HG LEMCVDARTFGNEARFIRRSCTPNAEVR HEIQDGTIHLYIYSIHSIPKGTEITIAFDFD YGNCKYKVDCACLKENPECPVLKRSSES MENINSGYETRRKKGKKDKDISKEKDT QNQNITLDCEGTTNKMKSPETKQRKLSP LRLSVSNNQEPDFIDDIEEKTPISNEVEM ESEEQIAERKRKMNWTVDGCGSWFWFR AEASVKALGLTLNVLFG
630	A	1826	5	SNLPEANVMSIVIPLGVDTAETSYLEMA AGSEPESVEASPVVVEKSNSYPHQLYTS SSHHSHSYNGFAPMRTHNRVLVPPPNTP APPPP\SVLISKNEVGHIYPLPNFDETSQC LPTISTS\EDGSYGTDVT\RCICGFTHDDG YMICCDKCSVWQHIDCMGIDRQHIPDTY LCERCQPRNLDKERAVLLQRRKRENMS DGDTSATESGDEVPVELYTAFQHTPTSIT LTASRVSKVNDKRRKKSGEKEQHISKCK KGSAPEIDPSSDGSNFGWETKIKAWMDR YEEANNNQYSEGVQREAQRIALRLGNG NDKKEMNKSDLNTNNLLFKPPVESHIQK NK\KILKSAK\DWPP\DALIIEYRGKFM\L R\EQFEANGYFF\KRP\YPFVLF\YSKF\HG LEMCVDARTFGNEARFIRRSCTPNAEVR HEIQDGTIHLYIYSIHSIPKGTEITIAFDFD

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				YGNCKYKVDCACLKENPECPVLKRSSESMENINSGYETRRKKGKKDKDISKEKDTQNQNITLDCEGTTNKMKSPETKQRKLSPLRLSVSNNQEPDFIDDIEEKTPISNEVEMESEEQIAERKRKMNWTVDGCGSWFWFRAEASVKALGLTLNVLFG
631	A	1197	878	LGHCNLSIYSYEIKTGVPSPRGMIYLLLCVCVCVCLPNHAIIVMLNRTEPYVPEGGAYLPEREPFIVPVEPERTAEYEDYGADEPAEEPPEPHRRWR\RALPHGPGQ
632	A	164	418	QIESSSLNVNKKDNHTKTPSEGHQHQRPKVDKSLKMRKKQRKKAENSKNQNASSPPKEHNSSPAKEQN*TENEFDELTEVGFRM
633	A	351	463	MESCSVTQVGVQWHNQLTATPPPRFKRLSS*DYRACI
634	A	88	328	RHKKRCLTYLTIK*MQIKTKMRYHFTHSRMFIVKKDSKCWQICGEIEILTHCWWKCTMVQSLWKIV*PFLKMMEKMRSEF
635	A	434	1142	AQGPYLARHEPLTASSLSRSAMSSEPPPPPQPPTHQASVGLLDTPWSRERSPSPLRGNVVPSPLPTRRTRTFSA*VTGSSGLGSTLNLQGAFLVSHPQPTAQPQVCLFGSPGHRPKNGQRHSESLASGKRSI
636	A	1931	233	FPARPTRPRTRGGVVVGAAAAGPRCGSARAQVAEGAVRNIQAEGHRQIPDLTWTTLLLLARPFNSPPLRSGLAPPPDLSLVLLQALTLDLRWLPHLSGAMATGADVRDILELGGPEGDAASGTISKKDIINPDKKKSKKSSETLTFKRPEGMHREVYALLYSDKKDAPPLLPSDTGQGYRTVKAKLGSKKVRPWKWMPFTNPARKDGAMFFHWRRAAEEGKDYPFARFNKTVQVPVYSEQEYQLYLHDDAWTKAETDHLFDLSRRFDLRFVVIHDRYDHQQFKKRSVEDLKERYYHICAKLANVRAVPGTDLKIPVFDAGHERRRKEQLERLYNRTPEQVAEEEYLLQELRKIEARKKEREKRSQDLQKLITAADTTAEQRRTERKAPKKKLPQKKEAEKPAVPETAGIKFPDFKSAGVTLRSQRMKLPSSVGQKKIKALEQMLLELGVELSPTPTEELVHMFNELRSDLVLLYELKQACANCEYELQMLRHRHEALARAGVLGGPATPASGPGPASAEPAVTEPGLGPDPKDTIIDVVGAPLTPNSRKRRESASSSSSVKKAKKP
637	A	1931	233	FPARPTRPRTRGGVVVGAAAAGPRCGSARAQVAEGAVRNIQAEGHRQIPDLTWTTLLLLARPFNSPPLRSGLAPPPDLSLVLLQALTLDLRWLPHLSGAMATGADVRDIL

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				ELGGPEGDAASGTISKKDIINPDKKKSKK SSETLTFKRPEGMHREVYALLYSDKKDA PPLLPSDTGQGYRTVKAKLGSKKVRPW KWMPFTNPARKDGAMFFHWRRAAEEG KDYPFARFNKTVQVPVYSEQEYQLYLH DDAWTKAETDHLFDLSRRFDLRFVVIHD RYDHQQFKKRSVEDLKERYYHICAKLA NVRAVPGTDLKIPVFDAGHERRRKEQLE RLYNRTPEQVAEEEYLLQELRKIEARKK EREKRSQDLQKLITAADTTAEQRRTERK APKKKLPQKKEAEKPAVPETAGIKFPDF KSAGVTLRSQRMKLPSSVGQKKIKALEQ MLLELGVELSPTPTEELVHMFNELRSDL VLLYELKQACANCEYELQMLRHRHEAL ARAGVLGGPATPASGPGPASAEPAVTEP GLGPDPKDTIIDVVGAPLTPNSRKRRESA SSSSSVKKAKKP
638	A	390	1258	LTRAPPPSTQVRPVPGPPAGAVVAAPGG ALASVSFDSRDSKMAAQSAPKVVLKSTT KMSLNERFTNMLKNKQPTPVNIRASMQ QQQQLASARNRRLAQQMENRPSVQAAL KLKQSLKQRLGKSNIQARLGRPIGALAR GAIGGRGLPIIQRGLPRGGLRGGRATRTL LRGGMSLRGQNLLRGGRAVAPRMGLRR GGVRGRGGPGRGGLGRGAMGRGGIGG RGRGMIGRGRGGFGGRGRGRGRGRGAL ARPVLTKEQLDNQLDAYMSKTKGHLDA ELDAYMAQTDPETND
639	A	1727	834	RPAEPDPGSGHPEPKSPTLSAVMSAEVK VTGQNQEQFLLLAKSAKGAALATLIHQ VLEAPGVVYVFGELLDMPNVRELAESDF ASTFRLLTVFAYGTYADYLAEARNLPPL TEAQKNKLRHLSVVTLAAKVKCIPYAV LLEALALRNVRQLEDLVIEAVYADVLRG SLDQRNQRLEVDYSIGRDIQRQDLSAIAR TLQEWCVGCEVVLSGIEEQVSRANQHK EQQLGLKQQIESEVANLKKTIKVTTAAA AAATSQDPEQHLTELREPAPGTNQRQPS KKASKGKGLRGSAKIWSKSN
640	A	56	443	RLHVVPMARYEEVSVSGFEEFHRAVEQ HNGKTIFAYFTGSKDAGGKSWCPDCVQ AEPVVREGLKHISEGCVFIYCQVGEKPY WKDPNNDFRKNLKVTAVPTLLKYGTPQ KLVESECLQANLVEMLFSED
641	A	320	43	LFYEEFSYATILTISSSEQKNSIKKSTFCLT LLPRLECSGAVMARCSLNLLGSGDPPAS ASQVAGTTGACHHTSLIFPFVLYVDSTG LTLKT
642	A	45	400	DPRVRTREWDSLNYENHIYLMPFIYINII

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				LGGTISLLGRLVYRSHLISSLLCLEGIILSLFIIATLITLNTHSLLTNIGPIAILVFAACEAAGGLSLLGSISNTYGLNYVHNLNLLQC
643	A	3828	2909	AECEEVRRKSELFNPVSLDCKLRQKAIAEVDVGTDKAQNSDPILDTSSLVPGCSSVDNIKSSQTSQNQGLGRPTLEGDEETSEVEYTVNKGPASSNRDRVPPSSEASEHHPRHRVSSQAEDTSSSFDNLFIDRLQRITIADQGEQQSEENASTKNLTGLSSGTEKKPHYMEVLEMRAKNPVPQLRKFKTNVLPFRQNDSSSHCQKSGSPISSEERRRRDKQHLDDITAARLLPLHHMPTQLLSIEESLALQKQQKQNYEEMQAKLAAQKLAERLNIKMRSYNPEGESSGRYREVRDEDDDWSSDEF
644	A	3	2185	RPYLGLLKMAALEEEFTLSSVVLSAGPEGLLGVEQSDKTDQFLVTDSGRTVILYKVSDQKPLGSWSVKQGQIITCPAVCNFQTGEYVVVHDNKVLRIWNNEDVNLDKVFKATLSAEVYRILSVQGTEPLVLFKEGAVRGLEALLADPQQKIETVISDEEVIKWTKFFVVFRHPVLIFITEKHGNYFAYVQMFNSRILTKYTLLLGQDENSVIKSFTASVDRKFISLMSLSSDGCIYETLIPIRPADPEKNQSLVKSLLLKAVVSGNARNGVALTALDQDHVAVLGSPLAASKECLSVWNIKFQTLQTSKELPQGTSGQLWYYGEHLFMLHGKSLTVIPYKCEVSSLAGALGKLKHSQDPGTHVVSHFVNWETPQGCGLGFQNSEQSRRILRRRKIEVSLQPEVPPSKQLLSTIMKDSEKHIEVEVRKFLALKQTPDFHTVIGDTVTGLLERCKAEPSFYPRNCLMQLIQTHVLSYSLCPDLMEIALKKKDVQLLQLCLQQFPDIPESVTCACLKIFLSIGDDSLQETDVNMESVFDYSINSVHDEKMEEQTEILQNGFNPEEDKCNNCDQELNKKPQDETKESTSCPVVQKRAALLNAILHSAYSETFLLPHLKDIPAQHITLFLKYLYFLYLKCSENATMTLPGIHPPTLNQIMDWICLLLDANFTVVVMMPEAKRLLINLYKLVKSQISVYSELNKIEVSFRELQKLNQEKNNRGLYSIEVLELF
645	A	118	473	WWPVLRRWTMKTFRWKVKPGMDVASVPSVRKVRFGDGYSQRAPAGLNANLKTYSVTLSVPREEATVLESFLEEHGGWKSFLWTPPYEWRQIKVTCAKWSSRVSMLRVEFSAEFEQVVN
646	A	138	632	QSVWVAQPPAVTADLQFLEGSGEKNAPTAVLRKRQEEMRPLDIDEVEAPEEVEVLEPEEDFEQFLLPVINEMREDIASLIREHGRAYLRTRSKLWEMDNMLIQIKTQVEASEE

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				SALNHVQHPSGEADERVSELCEKAEEKA KEIAKMAEMLVELVWRIERSESS
647	A	2224	1709	QASGLYSFLTLAPMTHLLLTATVTPSEQ NSSREPGWETAMAKDILGEAGLHFDEL NKLRVLDPEVTQQTIELKEECKDFVDKI GQFQKIVGGLIELVDQLAKEAENEKMK AIGARNLLKSIAKQREAQQQQLQALIAE KKMQLERYRVEYEALCKVEAEQNEFID QFIFQK
648	A	84	816	TGNKMQDPNADTEWNDILRKKGILPPK ESLKELEEEAEEEQRILQQSVVKTYEDM TLEELEDHEDEFNEEDERAIEMYRRRRL AEWKATKLKNKFGEVLEISGKDYVQEV TKAGEGLWVILHLYKQGIPLCALINQHL SGLARKFPDVKFIKAISTTCIPNYPDRNLP TIFVYLEGDIKAQFIGPLVFGGMNLTRDE LEWKLSESGAIMTDLEENPKKPIEDVLLS SVRRSVLMKRDSDSEGD
649	A	62	413	MTRQEELAAARAALHDLMTGKRVATV QKDGRRVEFTATSVSDLKKYIAELEVQT GMTQRRTGPAGFYV
650	A	196	339	RSYFPYSPGMFGDDCHQLCDCEGETFCH PKTEKCLCPRGRTGARCDAG
651	A	1280	906	LRMTTATRQEVLGLYRSIFRLARKWQA TSGQMEDTIKEKQYILNEARTLFRKNKN LTDTDLIKQCIDECTARIEIGLHYKIPYPR PIHLPPMGLTPLRGRGLRSQEKLRKLSKP VYLRSHDEVS
652	A	1359	39	CDAPLGGSIAGPFDQLEEILGWESAAAH SAELVAKGTVGKRKWGCAGASSSGSAL LPPCRELLMGHRFLRGLLTLLLPPPPLYT RHRMLGPESVPPPKRSRSKLMAPPRIGT HNGTFHCDEALACALLRLLPEYRDAEIV RTRDPEKLASCDIVVDVGGEYDPRRHRY DHHQRSFTETMSSLSPGKPWQTKLSSAG LIYLHFGHKLLAQLLGTSEEDSMVGTLY DKMYENFVEEVDAVDNGISQWAEGEPR YALTTTLSARVARLNPTWNHPDQDTEA GFKRAMDLVQEEFLQRLDFYQHSWLPA RALVEEALAQRFQVDPSGEIVELAKGAC PWKEHLYHLESGLSPPVAIFFVIYTDQAG QWRIQCVPKEPHSFQSRLPLPEPWRGLR DEALDQVSGIPGCIFVHASGFIGGHRTRE GALSMARATLAQRSYLPQIS
653	A	3	1652	QNAKVGSEAIQVLPAFPREVQRLPEPAK SCRQRQLEARKAAAEKKKIQKEKLSTPE KIKQEALELAGITSDPGLSLKGGLSQQGL KPSLKVEPQNHFSSFKYSGNAVVESYSV LGNCRPSDPYSMNSVYSYHSYYAQPSLT

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				SVNGFHSKYALPSFSYYGFPSSNPVFPSQ FLGPGAWGHSGSSGSFEKKPDLHALHNS LSPAYGGAEFAELPSQAVPTDAHHPTPH HQQPAYPGPKEYLLPKAPLLHSVSRDPS PFAQSSNCYNRSIKQEPVDPLTQAEPVPR DAGKMGKTPLSEVSQNGGPSHLWGQYS GGPSMSPKRTNGVGGSWGVFSSGESPAI VPDKLSSFGASCLAPSHFTDGQWGLFPG EGQQAASHSGGRLRGKPWSPCKFGNST SALAGPSLTEKPWALGAGDFNSALKGSP GFQDKLWNPMKGEEGRIPAAGASQLVF YQHKNLNQPNHGLALWEAKMKQLAER ARARQEEAARLGLGQQEAKLYGKKRK WGGTVVAEPQQKEKKGVVPTRQALAV PTDSAVTVSSYAYTKVTGPYSRWI
654	A	653	2	KYRWGFTVLARMVSIS*SRDLPVSASQS AGIIGVSHRARPRFYFLISLWHLLYFLIFF QS*Q*MKNLCLKKLPDDSNYCSCTTLFK/ TGNLSCLPDILCTKS*YSVS/CKSFANSYI QKKHDKM*KENGVR*GIINNLRLSIILILI LIS*LSYFNDFH*NSGTSLIVS*HYYLNKH DLHPNYFILLFCFSF/VRQGLRSVAQAGV QWCDLGSLEPPPPRF
655	A	415	0	SKSPPPKRKLFPSSPPKTFYPPPHSGVFPP FPP*NFFFSPRGLIFGGGFFQFFPPPKKRFF SKIPEAFLKVPPKRKKIVFATAPV/YFGPP REFFKGPPP
656	A	429	0	PSSSSPFFFFFPPPQKGIFFPPFFFCFPRVFS PPPFFIPPPRIFFLGPLKKKKTPPPPGGKNF FF*RAPPSP
657	A	222	140	GYWGCAIGKARQAAKTEIEKLQVIINA
658	A	30	250	YPMFQVQPVQHVYPAQVQYVEGGDAV YTNGAMCVLLPRATPGTLSLTPAFATW GGIKQIPLPYAFGDVFTKYC
659	A	1	707	HRPDGCPDIGPTGELSGSLKIPNRDSGIDS PSSSVAGENFPCEEGLEAGPSPTVLGAH AEMALDSQVPKVTPQEEADSDVGEEPD SENTPQKADKDAGLAQHSGPQKLLHIA QELLHTEETYVKRLHLLDQVFCTRLTDA GIPPEVIMGIFSNISSIHRFHGQFLLPELKT RITEEWDTNPRLGDILQKLAPFLKMYGE YVKNFDRAVGLVSTWTQRSPLFKDVVH SIQVRRRRG
660	A	379	3	MAEMQLAELRAEIKHFVSERKYDEELG KAARFSCDIEQLKAQIMLCGEIPHPKNN YSSRTPCSSLLPLLNAHAATSGKQSNFSR KSSTHNKPSEGKAANPKMVSSLPSTADP SHQTMTANKQNGE
661	A	379	3	MAEMQLAELRAEIKHFVSERKYDEELG

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				KAARFSCDIEQLKAQIMLCGEIPIHPKNN YSSRTPCSSLLPLLNAHAATSGKQSNFSR KSSTHNKPSEGKAANPKMVSSLPSTADP SHQTMTANKQNGE
662	A	3	783	TLQTLAPATVQTVAAPQVQQVPVSGWK GFREALVGAVNWSLVEDVVFMAAVQN PALTALTTPIQTAALQVPVRAAFSPTLLP PPLFLEVLVFRDVKQVPGGVKQLEPPKE GERRTT\HNIIEKRIRSS\INDKIIELKDLV MGTDAKVGAKQNCVLCSTISPLPVFASG A*RSPRLKVSEYKTELEVLFY*SILVSCQ DYKILKNIRDW*GLIRQ*PAETFDSFGYA LQLMGTPLSEAEKDACHSVVPDTGLCVP WSQLMA
663	A	346	28	YITPGICPSLSTMKAIVECAGGKVLSKQP SFRKLMEHKQNSVGRVESASESVIALYH IQRSVTEKLIFFSL\SLSEIILISCENDLHLC REYFARGIGTFPRLPV
664	A	558	175	ASLSCGGLHPVRASWLLCLTKQAWAME GAPPPASLPPCSLISDCCASNQRDSVGVG PSEPCAGYNLVVRRFLSRSEKRNIRVGV TRFSRCV/LSPLSLTQKGNSLTPCASQAR QCLALLRLAHGARTH
665	A	3	377	GRVRLL/NNWDVCANMFGTFFDTEDPG EFKMKYWGVTSFLQKGNDDHWIVDTD YDTYAGQYSCRLLNLDGTCADSYSFVFS RDPNGLPPEAQKIVRQRQEELCLARQYR LIVHNGYCDGRSKRNLL
666	A	355	250	QAGLELLTSDDLPASASKSAGTTGVSHR ARPTLVF
667	A	2	118	GNVLPVLYGEMRVGSRVVSQEISTADEG DGGQVVVIGR
668	A	3	339	TEQKSKSKQHRVSRRAQQRAESPESSAI ESTQSTPQKGRGRPSKTPSPSQPKKNVR VGRSKQAATKENDSSEEVNVFQGSSPVD DIPQEETEEEEVSTVNVRRRSAKRERR
669	A	2	120	NYRRRPRPPNAPSQDGKEAKAGEAPTEN PAPPTQQSSAE
670	A	2	120	NYRRRPRPPNAPSQDGKEAKAGEAPTEN PAPPTQQSSAE
671	A	3	284	KTENDHINLKVAGQDGSLVQFKIKRRTP LSKLMKAYCERQGLSMRHITLRFDGQSI NETDTPAQLEMEDEDTIDVFQQQTESVP DSILAGHSF
672	A	2	439	ERVLENVEHYQELKKMVQQSDLGQYVT FLRSFSDKQKISLLHSCTCVLYTPSNEHF GIVPLEAMYMQCPVIAVNSGGPLESIDH SVTGFLCEPDPVHFSEAIEKFIREPSLKAT MGLAGRARVKEKFSPEAFTEQLYRYVT

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				KLAG
673	A	1	631	KADFMRPRISFVDRCYIFSMLDNEVGAN PRGGHQCGSALSCDEGVPAAQGASSEKP HECTKCGKALCCRSDLRVHHGVHAGEK SSACSERGSGFREKLCPDKQGTHTKEKP ARDSRSGKTIFRKTRLCVPGTVHAGAKP YKCWECEKTSHKSRLIEHLRSHTGEKPC GCRECGKAFFQKSHLILRQRTHTGEKPC GCAECGKAAPRTPAS
674	A	3	242	VWPLPCRDADAQSEGINASARYPKNWV TTGDPAREFTMIHSAPLMLLADPDEFES VQLAQSWPLGAIVSLWRSPCRKMN
675	A	3	501	RGININKKDVHTETPCKGHQHQRPKVD KSMKMRKS*CKKAENSKNQNGSSPAKD YNSSPAREQNWMENWLDKLTEVGFIRW IITNSFKLKEHVLTQCKELRS/FEKKLEEL LTKIISLEKTISV*GEG*IQILDRSL*RVCV GYLRRSC
676	A	1	405	AAAAAADMERQEESLSARPALETEGLRF LHTTEGSLLATYGWNIVFSCILLYVVFQ KLSARLRALRHRQLDRAAAAVEPDVDV KLQEALAAARLKMQEELNAQVEKHKE KLKQLE*EKRRQTIEMWDSMQDGKSY
677	A	588	3	GGWLAFPLACPGRCHQPPAAYASAALL NLGRPEAPAHPQTAGSVRPAAVPPGC*G RTGGAGSPAGHSATGPAAPPPRGSTSSA PPARQAEAGLESSSAQVIRPPRSSKVLGL QA*ATAPGLPQLSEGDPGFPSGRVTWIG QSPASEDSWRPSAVSSLTYSPVSPARGC RSPS*KSGKT*KSRQPKRENPPGTAP
678	A	98	314	ILKAQATVGISINLDSPFTSPPLREEIMAN NFSLESHNISLTEHSSMPVENNITLERPSN VNLTCQFTTSGD
679	A	386	545	GNMKKQGISNFISYIYTHTHTHTHTHTYI YIYIHIYTYTHTYNFFFLERELHS
680	A	311	89	RRPECVWGRRGPAGEESEQTALIPLPRG LLTPPNWSGPPGLPPAASQPIILPGPGPSP NLLKSSTSSERSRGAK
681	A	860	84	PWVLTMNFSGGGRQEAAGSRGRRAPRP REQDRDVQLSKALSYALRHGALKLGLP MGADGFVPLGTLLQLPQFRGFSAEDVQ RVVDTNRKQRFALQLGDPSTGLLIRANQ GHSLQVPKLELMPLETPQALPPMLVHGT FWKHWPSILLKGLSCQGRTHIHLAPGLP GDPGIISGMRSRCEIAVFIDGPLALADGIP FFRSANGVILTPGNTDGFLLPKYFKEALQ LRPTRKPLSLAGDEETECQSSPKHSSRER RRIQQ
682	A	420	22	FDQPPGLKATRDEKDSLLNIARGKKNGE

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				KTRRVSSRKKPALKATSDEKDSFSNITR GKKDGEISRKVSSQKPPTLKGTSDEEDS VLGIARENKDGEKSRTVSSEKPPGLKAS SDEKENVLNIARVYCVRCRGS
683	A	1248	103	EALHLAHFGLTGFTQTHCARAWGGQTR KRRRVPRDPGGANCAGGGVSADTSGLA AAARGPPTLGFRERRSARSLDRIRPILVL MVDKKLVVVFGGTGAQGGSVARTLLE DGTFKVRVVTRNPRKKAAKELRLQGAE VVQGDQDDQVIMELALNGAYATFIVTN YWESCSQEQEVKQGKLLADLARRLGLH YVVYSGLENIKKLTAGRLAAAHFDGKG EVEEYFRDIGVPMTSVRLPCYFENLLSHF LPQKAPDGKSYLLSLPTGDVPMDGMSV SDLGPVVLSLLKMPEKYVGQNIGLSTCR HTAEEYAALLTKH\PARSCTMPR*LLRTT KSLAFPVPGTW/LNMFRFYALRPDRDIEL TLRLNPKALTLDQWLEQHKGDFNLL
684	A	148	2357	PNSMVVEHPEFLKAGKEPGLQIWRVEKF DLVPVPTNLYGDFFTGDAYVILKTVQLR NGNLQYDLHYWLGNECSQDESGAAAIF TVQLDDYLNGRAVQHREVQGFESATFL GYFKSGLKYKKGGVASGFKHVVPNEVV VQRLFQVKGRRVVRATEVPVSWESFNN GDCFILDLGNNIHQWCGSNSNRYERLKA TQVSKGIRDNERSGRARVHVSEEGTEPE AMLQVLGPKPALPAGTEDTAKEDAANR KLAKLYKVSNGAGTMSVSLVADENPFA QGALKSEDCFILDHGKDGKIFVWKGKQ ANTEERKAALKTASDFITKMDYPKQTQ VSVLPEGGETPLFKQFFKNWRDPDQTDG LGLSYLSSHIANVERVPFDAATLHTSTA MAAQHGMDDDGTGQKQIWRIEGSNKV PVDPATYGQFYGGDSYIILYNYRHGGRQ GQIIYNWQGAQSTQDEVAASAILTAQLD EELGGTPVQSRVVQGKEPAHLMSLFGG KPMIIYKGGTSREGGQTAPASTRLFQVR ANSAGATRAVEVLPKAGALNSNDAFVL KTPSAAYLWVGTGASEAEKTGAQELLR VLRAQPVQVAEGSEPDGFWEALGGKAA YRTSPRLKDKKMDAHPPRLFACSNKIGR FVIEEVPGELMQEDLATDDVMLLDTWD QVFVWVGKDSQEEEKTEALTSAKRYIET DPANRDRRTPITVVKQGFEPPSFVGWFL GWDDDY\WSVGPLGTGPMAGAGLP
685	A	361	570	NCFLVFSPARSTVGEFASMSSEECI*MHV VFLFPLTLCVVSL*ERQEAEKMFKGKRG AQLAKDIARRSKT
686	A	2484	1122	QAQPMGRVGGMAQPMGRAGAPKPMG

SEQ ID NO:	Method	Predicted beginning nucleotide location corresponding to first amino acid residue of peptide sequence	Predicted end nucleotide location corresponding to last amino acid residue of peptide sequence	Amino acid sequence (A=Alanine C=Cysteine, D=Aspartic Acid, E=Glutamic Acid, F=Phenylalanine, G=Glycine, H=Histidine, I=Isoleucine, K=Lysine, L=Leucine, M=Methionine, N=Asparagine, P=Proline, Q=Glutamine, R=Arginine, S=Serine, T=Threonine, V=Valine, W=Tryptophan, Y=Tyrosine, X=Unknown, *=Stop codon, /=possible nucleotide deletion, \=possible nucleotide insertion
				RAGSARRGRFKGCWSEGSPVHPVPAVL SWLLALLRCASTMLSLRVPLAPITDPQQ LQLSPLKGLSLVDKENTPPALSGTRVLA SKTARRIFQEPTEPKTKAAAPGVEDEPLL RENPRRFVIFPIEYHDIWQMYKKAEASF WTAEEVDLSKDIQHWESLKPEERYFISH VLAFFAASDGIVNENLVERFSQEVQITEA RCFYGFQIAMENIHSEMYSLLIDTYIKDP KEREFLFNAIETMPCVKKKADWALRWI GDKEATYGERVVAFAAVEGIFFSGSFASI FWLKKRGLMPGLTFSNELISRDEGLHCD FACLMFKHLVHKPSEERVREIIINAVRIE QEFLTEALPVKLIGMNCTLMKQYIEFVA DRLMLELGFSKVFRVENPFDFMENISLE GKTNFFEKRVGEYQRMGVMSSPTENSF TLDADF
687	A	25	219	RNMAAATLTSKLYSLLFRRTSTFALTIIV GVMFFERAFDQGADAIYDHVNEGKLW KHIKHKYENK
688	A	1	1034	QWRNSSHRCLLLHPPLRRLPPALPPPPPP CCCSKLLPPLGLKGFQMEHFDASLSTYF KALLGPRDTRVKGWFLLDNYIPTFICSVI YLLIVWLGPKYMRNKQPFSCRGILVVYN LGLTLLSLYMFCELVTGVWEGKYNFFC QGTRTAGESDMKIIRVLWWYYFSKLIEF MDTFFFILRKNNHQITVLHVYHHASMLN IWWFVMNWVPCGHSYFGATLNSFIHVL MYSYYGLSSVPSMRPYLWWKKYITQGQ LLQFVLTIIQTSCGVIWPCTFPLGWLYFQI GYMISLIALFTNFYIQTYNKKGASRRKD HLKDHQNGSMAAVNGHTNSFSPLENNV KPRKLRKD

TABLE 7

SEQ ID NO:	Position of end of signal in amino acid sequence	MaxS (maximum score)	MeanS (mean score)
1	29	0.962	0.801
5	17	0.989	0.926
15	33	0.933	0.780
20	39	0.899	0.608
23	18	0.992	0.867
25	32	0.928	0.768
26	37	0.979	0.808
27	48	0.977	0.661
28	18	0.965	0.805
29	16	0.972	0.860
30	21	0.948	0.785
31	23	0.954	0.824
32	18	0.966	0.896
33	15	0.989	0.956
39	16	0.988	0.971
61	37	0.979	0.808
62	21	0.990	0.901
73	48	0.977	0.661
86	18	0.965	0.805
87	16	0.949	0.749
88	16	0.949	0.749
89	16	0.949	0.749
91	16	0.972	0.860
102	21	0.948	0.785
103	19	0.893	0.604
104	28	0.945	0.556
106	17	0.965	0.869
107	16	0.980	0.957
108	22	0.969	0.917
115	23	0.954	0.824
126	18	0.966	0.896
138	38	0.901	0.607
139	33	0.976	0.754
140	27	0.966	0.909
147	38	0.901	0.607
149	22	0.929	0.838
162	32	0.949	0.704
168	27	0.966	0.909
171	42	0.895	0.609

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-172, or 345-516, a mature protein coding portion of SEQ ID NO: 1-172, or 345-516, an active domain coding portion of SEQ ID NO: 1-172, or 345-516, and complementary sequences thereof.

2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.

3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.

4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.

5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.

6. A vector comprising the polynucleotide of claim 1.

7. An expression vector comprising the polynucleotide of claim 1.

8. A host cell genetically engineered to comprise the polynucleotide of claim 1.

9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.

10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:
   (a) a polypeptide encoded by any one of the polynucleotides of claim 1; and
   (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-172, or 345-516.

11. A composition comprising the polypeptide of claim 10 and a carrier.

12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and

b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,

a) culturing a host cell comprising a polynucleotide sequence selected from SEQ ID NO: 1-172, or 345-516, a mature protein coding portion of SEQ ID NO: 1-172, or 345-516, an active domain coding portion of SEQ ID NO: 1-172, or 345-516, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-172, or 345-516, under conditions sufficient to express the polypeptide in said cell; and

b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of any one of the polypeptides SEQ ID NO: 173-344, or 517-688, the mature protein portion thereof, or the active domain thereof.

21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprising the sequence information of at least one of SEQ ID NO: 1-172, or 345-516.

23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

28. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.